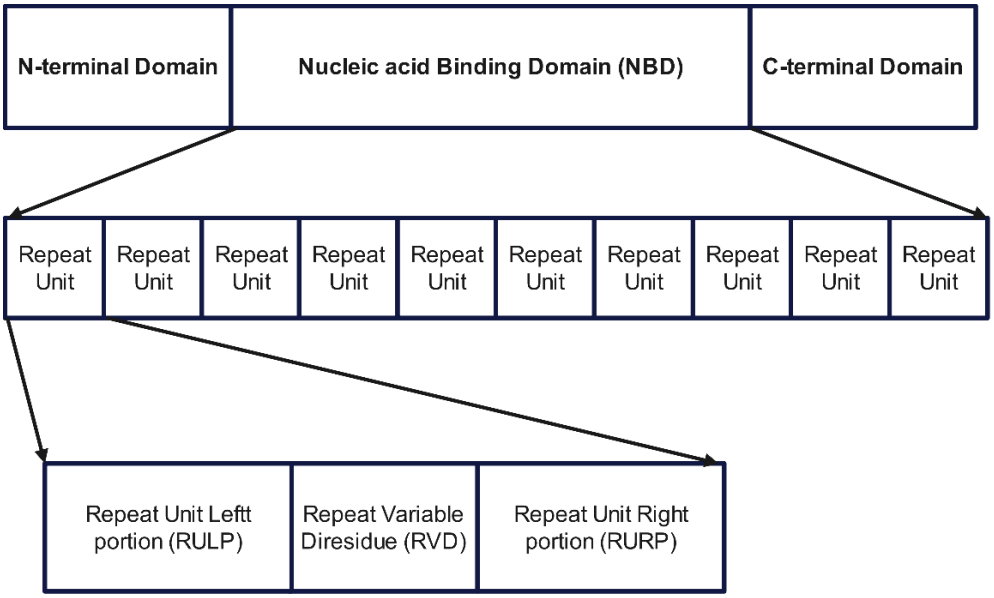
Title of invention:	Programmable Nucleic Acid-Binding polypeptides from bacterial proteins.	
Inventor: **(and author)**	**First Name**	**Last name**
	Rojopitiavana	Solofondrazaintsoanirina
City and Country of residence:	Antananarivo, Madagascar (MG)	
CONTACT:	solofo.rojo@gmail.com rojo.solofo@yahoo.fr	
Publication date:	Feb. 8, 2023	

Disclaimer: The instructions set forth in this document are provided for informational purposes only. The inventor assumes no responsibility or liability for any damage or loss that may occur as a result of following these instructions. The reader assumes all risk and responsibility for any consequences arising from the use or application of the disclosed inventions.

Digitally signed by Rojopitiavana Solofondrazaintsoanirina
Reason: I am the author of this document
Date: 2023-02-08 13:24:50

ABSTRACT

The present invention provides Modular Nucleic Acid-Binding Proteins (MNABPs) comprising a nucleic acid-binding domain (NBD), wherein the NBD comprises a plurality of repeat units derived from proteins sourced from *Legionellales bacterium, Gammaproteobacteria bacterium, Candidatus Symchoanobacter obligatus, Apophysomyces sp. BC1034, Apophysomyces sp. BC1015, Burkholderiales bacterium, Burkholderia metallica, Legionella sp. W10-070, Legionella yabuuchiae, Pseudomonas quercus, and/or Pseudomonas sp. LY10J*.



FIGURES

FIGURE 1

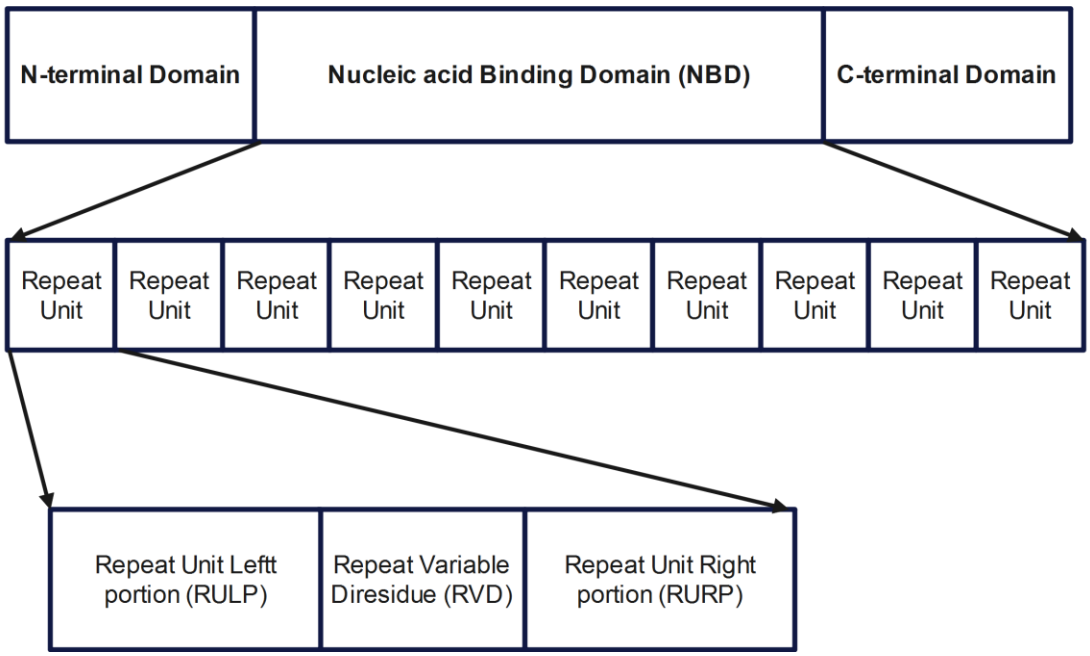


FIGURE 2

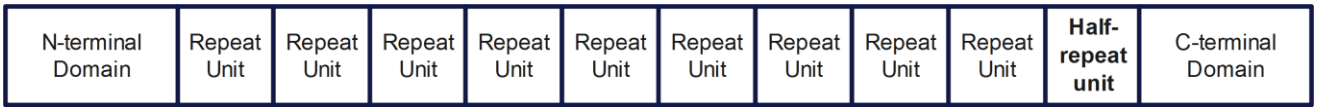


FIGURE 3

FIGURE 4

FIGURE 5

FIGURE 6

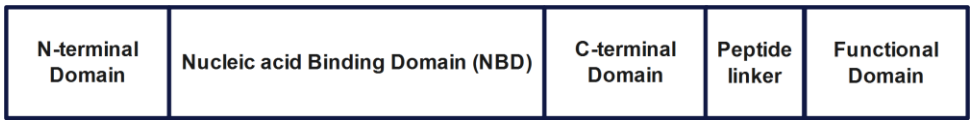


FIGURE 7

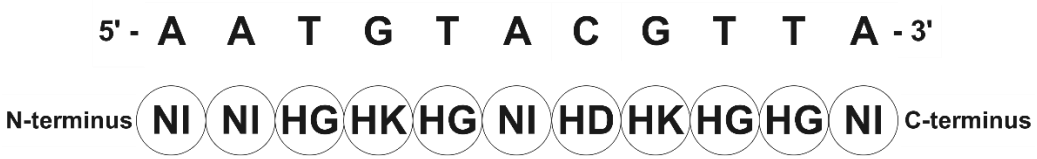


FIGURE 8

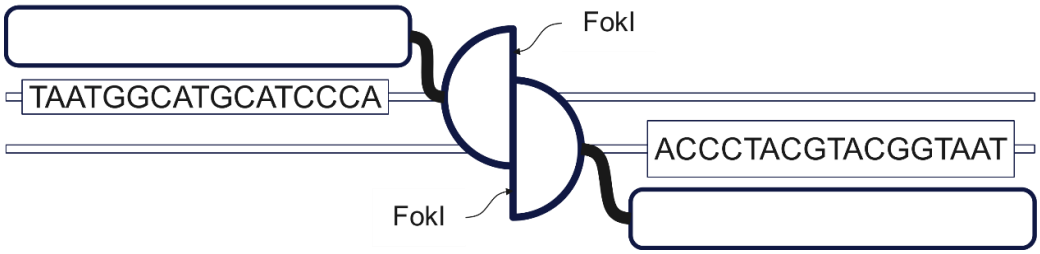
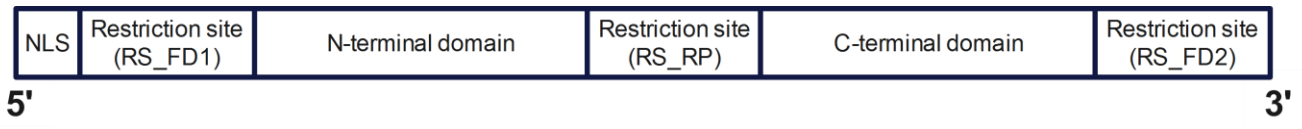


FIGURE 9



BRIEF DESCRIPTION OF THE FIGURES

[0001] FIGURE 1: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein. Ordered from N-terminus to C-terminus are: the N-terminal domain, the Nucleic acid-Binding Domain (NBD), and the C-terminal domain. The herein NBD comprises a plurality of repeat units, wherein each one repeat unit comprises independently: a Repeat Unit Left Portion (RULP), a Repeat Variable Di-residue (RVD), and a Repeat Unit Right Portion (RURP)

[0002] FIGURE 2: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein, wherein the last repeat unit of its Nucleic acid-Binding Domain (NBD) is a half-repeat unit.

[0003] FIGURE 3: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein, wherein its C-terminus is fused to the N-terminus of a Functional Domain.

[0004] FIGURE 4: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein, wherein its N-terminus is fused to the C-terminus of a Functional Domain.

[0005] FIGURE 5: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein, wherein both its N-terminus and C-terminus are fused to Functional Domains.

[0006] FIGURE 6: Schematic representation of an embodiment of a Modular Nucleic Acid-Binding Protein (MNABP) disclosed herein, wherein a peptide linker connect its C-terminus to the N-terminus of a Functional Domain.

[0007] FIGURE 7: Illustration of a correspondence between the nucleotide bases of a target sequence and the RVDs of an RVD sequence.

[0008] FIGURE 8: Illustration of two Modular Nucleic Acid-Binding ENdonucleases (MNABENs) engaging their respective target nucleic acid sequences, wherein their FokI domains dimerize.

[0009] FIGURE 9: Graphical representation of an embodiment of a polynucleotide sequence (a DNA) comprising, from 5’ to 3’: a Nuclear Localization Signal (NLS), a Restriction Site (RS_FD1), a sequence encoding a N-terminal domain, a Restriction Site (RS_RP), a sequence encoding a C-terminal domain, and a Restriction Site (RS_FD2).

BACKGROUND

[0010] Transcription activator-like effectors (TALEs) are proteins synthesized by plant pathogenic bacteria of the Xanthomonas genus that function as DNA-binding proteins. They act as transcriptional activators, bind to specific DNA sequences, and activate gene expression, leading to changes in the host plant's physiology that benefit the bacterium. TALEs are unique in that the DNA-binding specificity of each TALE is determined by an array of domains of 33-35 amino acids repeats, making them useful as tools in genome engineering and synthetic biology (Boch et al., 2009). Each base on the same strand of target DNA interacts with a single TALE repeat at the 12th and 13th positions, known as the Repeat Variable Di-residue (RVD). The specificity of the interaction of each repeat with a given base of the target DNA is driven by the sequence of the RVD.

[0011] A TALE protein that binds a specific DNA sequence can be obtained by assembling an array or repeats that differ in their RVDs in respect of the nucleotide bases they recognize in the targeted sequence.

[0012] An example of Transcription activator-like effectors (TALEs) proteins is AvrBs3. Other proteins with sequence similarity to the TALEs were discovered and characterized for their DNA-binding activities.

[0013] Brg11 (Cunnac et al., 2004), a protein from the plant pathogen *Ralstonia solanacearum*, shares 40% homology with the protein AvrBs3 (Schornack et al., 2006) and comprises a core tandem repeats of 35 amino acids. Brg11 was found to recognize DNA in a manner analogous to that of TALEs (de Lange et al., 2013).

[0014] EAV36, E5AW43, E5AW45 and E5AW46 (values are uniprot accessions) are proteins from the bacteria endosymbiont *Burkholderia rhizoxinica*, termed Bat proteins, which possess highly polymorphic tandemly arranged 31-34 amino acids repeats (de Lange et al., 2014). The repeat arrays of bat proteins were discovered to mediate base-per-base specific DNA binding using the same rules as TALEs.

[0015] The following proteins (values in brackets are protein NCBI accessions): [WP_058473422.1] from *Legionella quateirensis*, [WP_058451450.1] from *Legionella maceachernii*, [AXA34194.1] from *Francisella adeliensis*, [WP_173262489.1, WP_173262634.1, WP_173263036.1, WP_173263118.1, WP_254367233.1] from *Paraburkholderia sp. NMBU_R16*, [OGV28801.1] from a *Legionellales*, [OXJ06552.1, WP_089477027.1] from *Burkholderia sp. AU6039*, [SIT73265.1] from *Burkholderia sp. b13*, and [SIT71710.1, SIT64975.1, SIT64981.1] from *Burkholderia sp. b14*, possess 31-34 amino acids long tandem repeats, each one of which mediates the recognition of a base in a target nucleic acid in a manner similar to TALE repeats (Urnov F et al., 2018).

[0016] Many TALE and TALE-like repeats, herein referred to as Repeat Units (RUs), are well known in the prior art. Examples of patents that disclose RUs' methods of identification, their isolation, their amino acid sequences, their design, and/or assembly to form multi-domains nucleic acid binding proteins, and methods of uses thereof for genome editing or gene regulation, include: US9017967B2 (Modular DNA-binding domains and methods of use), WO2011072246A2 (Tal effector-mediated DNA modification), WO2014018601A2 (New modular base-specific nucleic acid binding domains from burkholderia rhizoxinica proteins), WO2014167058A1 (RALEN-mediated genetic modification techniques), WO2019204643A2 (Animal pathogen-derived polypeptides and uses thereof for genetic engineering).

DESCRIPTION

[0017] The present invention concerns novels Modular Nucleic Acid-Binding Proteins (MNABPs) derived from the following proteins (the values are protein NCBI accessions): WP_267874466.1, WP_229632017.1, WP_218585518.1, WP_218585522.1, WP_195755884.1, WP_195755899.1, WP_195755907.1, WP_195755908.1, WP_195755912.1, WP_178089108.1, WP_178089118.1, WP_178089119.1, WP_168083840.1, WP_168086212.1, WP_226244440.1, WP_133129496.1, WP_133133554.1, KAG0189736.1, KAG0162681.1, TLY47364.1, TAK77069.1, TAK78877.1, WP_258568932.1, OGV35086.1, OGV33042.1, and MBW8829317.1. By "modular" is meant that the MNABPs are assembled or modified to bind a desired nucleic acid sequence or a target nucleic acid sequence.

[0018] Relevant information concerning a protein, for example WP_267874466.1, can be found by using or modifying the following link: https://www.ncbi.nlm.nih.gov/protein/WP_267874466.1 (you may replace WP_267874466.1 by any NCBI accession of your choice).

[0019] WP_267874466.1, WP_229632017.1, WP_218585518.1, WP_218585522.1, WP_195755884.1, WP_195755899.1, WP_195755907.1, WP_195755908.1, WP_195755912.1, WP_178089108.1, WP_178089118.1, WP_178089119.1, WP_168083840.1, and WP_168086212.1 are proteins derived from *Pseudomonas quercus* and/or *Pseudomonas sp. LY10J*.

[0020] WP_226244440.1 is a protein derived from *Burkholderia metallica*.

[0021] WP_133129496.1 is a protein derived from *Legionella yabuuchiae*.

[0022] WP_133133554.1 is a protein derived from *Legionella sp. W10-070*.

[0023] KAG0189736.1 is a protein derived from *Apophysomyces sp. BC1034*.

[0024] KAG0162681.1 is a protein derived from *Apophysomyces sp. BC1015*.

[0025] TLY47364.1, TAK77069.1, and TAK78877.1 are proteins derived from Gammaproteobacteria bacteria.

[0026] WP_258568932.1 is a protein derived from *Candidatus Symchoanobacter obligatus*.

[0027] OGV35086.1, and OGV33042.1 are proteins derived from *Legionellales* bacteria.

[0028] MBW8829317.1 is a protein derived from a *Burkholderiales* bacterium.

[0029] A Modular Nucleic Acid-Binding Protein (MNABP) of the present invention binds to a target nucleic acid or a desired nucleic acid. The nucleic acid can be a double-stranded DNA, a double-stranded RNA, a single-stranded DNAs that forms a duplex, a single-stranded RNA that forms a duplex, or a hybrid RNA-DNA duplex.

[0030] The Modular Nucleic Acid-Binding Protein (MNABP) consists, ordered from N-terminus to C-terminus, an N-terminal domain, a Nucleic acid-Binding Domain (NBD), and a C-terminal domain.

[0031] In certain aspects, the Nucleic acid-Binding Domain (NBD) comprises a plurality of repeat units.

[0032] In certain aspects, the Nucleic acid-Binding Domain (NBD) comprises, ordered from the N-terminus to the C-terminus, a plurality of repeat units, and a half-repeat unit.

[0033] Each one repeat unit of the plurality of repeat units comprises independently, ordered from the N-terminus to the C-terminus: (a) a Repeat Unit Left Portion (RULP); (b) a Repeat Variable Di-residue (RVD); (c) a Repeat Unit Right Portion (RURP).

[0034] A repeat unit is about 31 to 35 amino acids in length, wherein two critical amino acids at positions 12 and 13 - the Repeat Variable Di-residue (RVD) - determine the specificity of interaction with the nucleotide base of the target nucleic acid.

[0035] In some instances, a repeat unit is selected from any one of the sequences provided herein in Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10, or Table 11, in respect of the correspondence of the RVD of the selected repeat unit to a given nucleic acid base set forth in Table 12.

[0036] TABLE 1: Repeat units derived from WP_267874466.1, WP_229632017.1, WP_218585518.1, WP_218585522.1, WP_195755884.1, WP_195755899.1, WP_195755907.1, WP_195755908.1, WP_195755912.1, WP_178089108.1, WP_178089118.1, WP_178089119.1, WP_168083840.1, and WP_168086212.1

SEQ ID	Repeat Unit	RVD
RU_001	FGNDNLVKVAAHDGGAQALQALLDKGPALRQAG	HD
RU_002	FGNDNLVKVAAHDGGAQALQALLDKGPTLRQAG	HD
RU_003	FGNDNLVKVAAHDGGAQALQALLDKGTALRQAG	HD
RU_004	FGNDNLVKVAAHDGGAQALQALLDRGPALRQAG	HD
RU_005	FGNDNLVKVAANGGGAQALQALLDKGPALRQAG	NG
RU_006	FGNDNLVKVAANIGGAQALQALLDKGPALRQAG	NI
RU_007	FGNDNLVKVAANNGGQHALQALLDKGPALRQAG	NN
RU_008	FGNDNLVKVAANNGGQQALQALLDKGPALRNAG	NN
RU_009	FGNDNLVKVAANNGGQQALQALLDRGPALRQAG	NN
RU_010	FGNDNLVKVAANNGSQHALQALLDKGPALRQAG	NN
RU_011	FGNDNLVKVAANNGSQQALQALLDKGPALRQAG	NN
RU_012	FGPDNLVKVAAHDGGAQALQALLDKGPALRQAG	HD
RU_013	FGPDNLVKVAAHDGSAQALQALLDKGPALRQAG	HD
RU_014	FGPDNLVKVAAHNGGQQALQALLDKGPALRQAG	HN
RU_015	FGPDNLVKVAANGGGAQALQALLDKGPTLRQAG	NG
RU_016	FGPDNLVKVAANIGGAQALQALLDKGPALLQAG	NI
RU_017	FGPDNLVKVAANKGGQQALQALLDKGPALRQAG	NK
RU_018	FGPDNLVKVAANNGGQQALQALLDKGPALRQAG	NN
RU_019	FGPDSLVKVAANNGGQHALQALLDKGPALRQAG	NN
RU_020	FSADNLVRIAANNGGQQALQALLDKGPALRNAG	NN
RU_021	FSADNLVRIAANNGGQQALQALLDKGPALRQAG	NN
RU_022	FSLDNLVKVAANIGGAQALQALLDKGPALRQAG	NI
RU_023	FSNDNLIKVAANIGGTQALQALLDKGPALRQAG	NI
RU_024	FSNDNLMRIAAHDGGAQALQALLDKGPALRQAG	HD
RU_025	FSNDNLVKVAAHDGGAQALQALLDKGPALRNAG	HD
RU_026	FSNDNLVKVAAHDGGAQALQALLDKGPALRQAG	HD

RU_027	FSNDNLVKVAAHDGGAQALQALLDKGSALRQAG	HD
RU_028	FSNDNLVKVAAHDGGAQALQTLLDKGPALRQAG	HD
RU_029	FSNDNLVKVAAHNGGQQALQALLDKGPALRNAG	HN
RU_030	FSNDNLVKVAANGGGAHALQALLDKGPALRQAG	NG
RU_031	FSNDNLVKVAANGGGAQALQALLDKGPALRQAG	NG
RU_032	FSNDNLVKVAANIGGAQALQALLDKGPALRQAG	NI
RU_033	FSNDNLVKVAANIGGTQALQALLDKGPALRQAG	NI
RU_034	FSNDNLVKVAANNGAQQALQALLDKGPALRQAG	NN
RU_035	FSNDNLVKVAANNGGQQALQALLDKGPALRQAG	NN
RU_036	FSNDNLVKVAANNGGQQALQTLLDKGPALRQAG	NN
RU_037	FSNDNLVKVAANTGGAQALQALLDKGPALRQAG	NT
RU_038	FSPDNLIKVAAYVGGAQALQALLDKSPALRQAG	YV
RU_039	FSPDNLVKVAAHDGGAQALQALLDKGPALRQAG	HD
RU_040	FSPDNLVKVAANGGGAHALQALLDKGPALRQAG	NG
RU_041	FSPDNLVKVAANNGSQQALQALLDKSPALRQAG	NN
RU_042	GGREQVIKIAAHHGGQQALQALLDKGPALRNAG	HH
RU_043	GGREQVIKIAAHHGGQQALQALLDKGPALRQAG	HH
RU_044	GGREQVIKIAANNGGKQALQALLDKSPALRQAG	NN
RU_045	GSREQVIKIAANHGGQQALQALLDKGPALRNAG	NH
RU_046	GSREQVIKIAANKGGQQALQALLDKSPALRQAG	NK

[0037] TABLE 2: Repeat units derived from WP_226244440.1

SEQ ID	Repeat Unit	RVD
RU_047	TKADIVKIASNSSGGMQALQAVINLHSELTKIG	SS
RU_048	LSNNNIVNIAANNGGSQALRAVFTHHPALIQAG	NN
RU_049	FSNDQIAKIAGNYGGAQTVQAVIDLYHLLTNAG	NY
RU_050	LNNKNIVKIAGNSGGAQALRAVVTHHPALIEAG	NS
RU_051	FSNDHVVKIGGNRGGAQALQAVANLHSSLEVAG	NR
RU_052	FGNNGIVRIAGNIGGAQALRAVITHGSALVQRG	NI
RU_053	FSNDDIVGIAGNNGGAQALQAVITHYPALIQAG	NN

[0038] TABLE 3: Repeat units derived from WP_133129496.1

SEQ ID	Repeat Unit	RVD
RU_054	FTGEQILKIVAHDGGSKNLNAVLLHFKALRALK	HD
RU_055	FNCKDIVKIVGHGGGSKNLNAVLAHSEALCALQ	HG
RU_056	FQVEDIVKILAHQGGSKNLNAVLAHSEALLALQ	HQ
RU_057	FTGQDILKMVGHDGGSKNLNAVLEHFEALRALQ	HD
RU_058	FTGEDIVKIVGHIGGSKNLGAVLVNFKTLRDLQ	HI

[0039] TABLE 4: Repeat units derived from WP_133133554.1

SEQ ID	Repeat Unit	RVD
RU_059	FNPEQIVKMVSHDGGSRNLDAVKNNVEALDKLG	HD
RU_060	FDSNQIVKMVSHGGGSKNLDAVKNNLEALDKLG	HG
RU_061	FDSNQIVKMVSHGGGSKNLDAVKNNAEALDKLG	HG
RU_062	FDSNQIVKMVSHGGGSKNLDAIKNNLEALDKLG	HG
RU_063	FDSNQIVKMVSHIGGSKNLDAVKNNVEILKALD	HI
RU_064	FNPEQIVKMVSHDGGSRNLDAVKNNVEILKALD	HD
RU_065	FNPEQIVKMVSHDGGSKNLDAVKNNLEILKALD	HD
RU_066	FNPEQIVKMVSHDGGSKNLDAVKNNLEALDKLG	HD
RU_067	FDSNQIVKMVSHGGGSRNLDAVKNNLEALDKLG	HG
RU_068	FDSNQIVKMVSHGGGSRNLDAVKNNADILKALE	HG
RU_069	FNPEQIVKMVSHIGGSRNLDAVKNNVEALDKLG	HI

[0040] TABLE 5: Repeat units derived from KAG0189736.1

SEQ ID	Repeat Unit	RVD
RU_070	FSKQEAVAIASNHGGSQALNTVLATHATLTAAG	NH
RU_071	FTHQQIVAIASKGGGSQALNTVLATHAALTAAG	KG
RU_072	FTHQQIVAIASNHGGSQALDKVLATHAPLTAAG	NH
RU_073	FTHRQIVGIASNNGGSQALDTVLVRYAPLRDAG	NN
RU_074	FKHEQIVGIASNIGGSQALDKVLATHAQLTAVG	NI
RU_075	FKHEQIVAIASKGGGSQALDKVLVKYAPLTAAG	KG
RU_076	FTHQQIVAIASNKGGSQALDTVLATHAQLTTAG	NK

[0041] TABLE 6: Repeat units derived from KAG0162681.1

SEQ ID	Repeat Unit	RVD
RU_077	FSKQEAVAIASNIGGSQALDKVLATHAQLTAVG	NI
RU_078	FKHEQIVAIASKGGGSQALDKVLVKYAPLTAAG	KG
RU_079	FTHQQIVAIASNKGGSQALDKVLATHAQLTTAG	NK

[0042] TABLE 7: Repeat units derived from TLY47364.1

SEQ ID	Repeat Unit	RVD
RU_080	YTQNQLNKIASNRGGSKTLNTLLEKAPQLLTLG	NR
RU_081	YKVEQIVKVAANSGGSKTLNTLLEKTPQLLALG	NS
RU_082	YKDEQLIKVAANSGGSKTLNTLLEKTPQLLALG	NS
RU_083	YKDEQLIKVAANSGGSKTLNTLLEKTPQLLTLG	NS
RU_084	YKDEQIVKVAANGGGSQALTTLLEKTPQLLILG	NG
RU_085	YKADQLIKAAANSGGSQALNTLLEKTPQLLTLG	NS

[0043] TABLE 8: Repeat units derived from TAK77069.1, and TAK78877.1

SEQ ID	Repeat Unit	RVD
RU_086	FTAEQMVKMVSPRGGSKNLEAIKNNYDALKELG	PR
RU_087	FTAEQMVKMVSHIGGSKNLEAIRYGSDVLKYLG	HI
RU_088	FTSEQLVDMVSYDGGSKNLEELKMSYYVLKDLG	YD
RU_089	FTVEQMVNMVSHNGGSKNLEAIRYSSDALKYLG	HN
RU_090	FTSEQMVNMVSHNGGSKNLEAIRYSYHVLKELG	HN
RU_091	FTTEQMVKMVKHSGGSKNLEAIKNNYDALKALG	HS
RU_092	FTAERMVKMASHIGGSKNLEIIKNNYDALKELG	HI
RU_093	FTAEQMVKMVNHSGGSRNLEAIKNNYDALKALG	HS
RU_094	FTAEQMVKMASNIGGSKNLEIIKNNYDVLKESG	NI

[0044] TABLE 9: Repeat units derived from WP_258568932.1

SEQ ID	Repeat Unit	RVD
RU_095	YSTADITRIAAHNGGSKNLEAVNLKHTELISLG	HN
RU_096	FNAIQIVSMVSHGGGSKNLQAVTDNNEALKDLS	HG

RU_097	FTAKQIVSIVSHDGGSKNLQAVTENNEALKDLG	HD
RU_098	FNAVQVVRMVSHKGGSKNLQAVTENHEALLNLS	HK
RU_099	FTAEQIVRMASHKGGSKNLQAVTENHEALLNLS	HK
RU_100	FTAEQIVSMVSHGGGSKNLQVVTDNNEALKDLG	HG
RU_101	FNAVQVVRMVSHKGGSKNLQAVTENNEALKGLG	HK
RU_102	FTAVQVVRMVSHSGGSKNLQAVTDNNEALKGLG	HS
RU_103	FTAKQIVRMVSHDGGSKNLQAITDNNEALLNLG	HD
RU_104	FTAAQIVSMVSHIGGSKNLQAVTDNNEALKGLG	HI

[0045] TABLE 10: Repeat units derived from OGV35086.1 and OGV33042.1

SEQ ID	Repeat Unit	RVD
RU_105	FPREEIGKIAGNNGGSHNLKAVLTHTQALINLG	NN
RU_106	FPREEIGKIAGHIGGSHNLEAVLTHARALIDLG	HI
RU_107	FPREEIGKIVGHDGGSRNLEAVLTHARALIDLG	HD
RU_108	FPCNEIGKIVGHGGGSRNLKAVLTHARALIDLG	HG
RU_109	FLREEIGKIAGHGGGSRNLKAVLTHARALIYLG	HG
RU_110	FPCEEIGKIAGHIGGSHNLEAVRTHVQALINLG	HI
RU_111	FPREEIGKIAGHGGGSHNLEAVLTHARALVDLG	HG
RU_112	FPREEIGKIAGHGGGSHNLEAVLTHAQALIHLG	HG
RU_113	FQREEIGKIAGHDGGSRNLEAVLTHAQALINLG	HD
RU_114	FPCEEIGVIAGNKGGSRNLDAVLTHARSLIDLG	NK
RU_115	FPHEEIGKIAGHIGGSRNLKAVLTHAQALIDLG	HI
RU_116	FSREEISKIAGHGGGSHNLEAVLKHFNVLEKLG	HG
RU_117	FTHAELVKIARNNGGSRNLKAVHVNAQALIDLG	NN
RU_118	FPREEVGKIAGHDGGSLNLEAMLTHARALIDLG	HD
RU_119	FQHEEICQIARHDGGSRNLKAVLTDAQSLIDLG	HD
RU_120	FPREEISKIAGNNGGSHNLAAVLKHVQTLIDLG	NN
RU_121	FPREEISKIAGHGGGSHNLAAVLKHVQTLIDLG	HG
RU_122	FQREEIGKIAGHGGGSLNLQAVLTNAQALIDLG	HG
RU_123	FSREEIGKIAGHDGGSRNLEAVNKHVQTLIDLG	HD
RU_124	FQHEEISKIAGHRGGSLNLQAVLTNAQALIDLG	HR
RU_125	FPREDIGKIAGRDGGSCNLEAMLKHFSILQKLG	RD

[0046] TABLE 11: Repeat units derived from MBW8829317.1

SEQ ID	Repeat Unit	RVD
RU_126	FSRTEIVSIASKGGGSQALGKVLATLERLKVAG	KG
RU_127	FEHKHIVAIAANIGGSQALDKVLDTHERLKNAG	NI
RU_128	FEHKHIVAIASKGGASQALDKVLSTHEQLKEAG	KG
RU_129	FEVNQIAAIATHKGGSRALDKVLAAHKQTKAGR	HK
RU_130	FDHEEIVNIASNDGGSQALAKVLATHDRLRSAG	ND
RU_131	FEHEHIVAIAAEIGGKQALEKVLSKHEQFKDAG	EI

[0047] TABLE 12: correspondence between RVD sequences and Nucleotide bases

RVD	Base
EI	A
HI	A
NI	A
NS	A, G, T, or C
NN	A, or G
HD	C
ND	C
RD	C
HH	G
HK	G
HN	G
NK	G
HG	T
KG	T
NG	T

[0048] In some instances, a repeat unit is assembled by: (a) selecting one pair of RULP and RURP from any one of Table 13, Table 14, Table 15, Table 16, Table 17, Table 18, Table 19, Table 20, Table 21, or Table 22; (b) selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes; (c) concatenating together each part in the following order: RULP, RVD, RURP.

[0049] TABLE 13: Pairs of RULP,RURP derived from WP_267874466.1, WP_229632017.1, WP_218585518.1, WP_218585522.1, WP_195755884.1, WP_195755899.1, WP_195755907.1, WP_195755908.1, WP_195755912.1, WP_178089108.1, WP_178089118.1, WP_178089119.1, WP_168083840.1, and WP_168086212.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_001	FGNDNLVKVAA	GGAQALQALLDKGPALRQAG
PAIR_002	FGNDNLVKVAA	GGAQALQALLDKGPTLRQAG
PAIR_003	FGNDNLVKVAA	GGAQALQALLDKGTALRQAG
PAIR_004	FGNDNLVKVAA	GGAQALQALLDRGPALRQAG
PAIR_005	FGNDNLVKVAA	GGQHALQALLDKGPALRQAG
PAIR_006	FGNDNLVKVAA	GGQQALQALLDKGPALRNAG
PAIR_007	FGNDNLVKVAA	GGQQALQALLDRGPALRQAG
PAIR_008	FGNDNLVKVAA	GSQHALQALLDKGPALRQAG
PAIR_009	FGNDNLVKVAA	GSQQALQALLDKGPALRQAG
PAIR_010	FGPDNLVKVAA	GGAQALQALLDKGPALRQAG
PAIR_011	FGPDNLVKVAA	GSAQALQALLDKGPALRQAG
PAIR_012	FGPDNLVKVAA	GGQQALQALLDKGPALRQAG
PAIR_013	FGPDNLVKVAA	GGAQALQALLDKGPTLRQAG
PAIR_014	FGPDNLVKVAA	GGAQALQALLDKGPALLQAG
PAIR_015	FGPDSLVKVAA	GGQHALQALLDKGPALRQAG
PAIR_016	FSADNLVRIAA	GGQQALQALLDKGPALRNAG
PAIR_017	FSADNLVRIAA	GGQQALQALLDKGPALRQAG
PAIR_018	FSLDNLVKVAA	GGAQALQALLDKGPALRQAG
PAIR_019	FSNDNLIKVAA	GGTQALQALLDKGPALRQAG
PAIR_020	FSNDNLMRIAA	GGAQALQALLDKGPALRQAG
PAIR_021	FSNDNLVKVAA	GGAQALQALLDKGPALRNAG
PAIR_022	FSNDNLVKVAA	GGAQALQALLDKGPALRQAG
PAIR_023	FSNDNLVKVAA	GGAQALQALLDKGSALRQAG
PAIR_024	FSNDNLVKVAA	GGAQALQTLLDKGPALRQAG
PAIR_025	FSNDNLVKVAA	GGQQALQALLDKGPALRNAG
PAIR_026	FSNDNLVKVAA	GGAHALQALLDKGPALRQAG
PAIR_027	FSNDNLVKVAA	GGTQALQALLDKGPALRQAG
PAIR_028	FSNDNLVKVAA	GAQQALQALLDKGPALRQAG
PAIR_029	FSNDNLVKVAA	GGQQALQALLDKGPALRQAG
PAIR_030	FSNDNLVKVAA	GGQQALQTLLDKGPALRQAG

PAIR_031	FSPDNLIKVAA	GGAQALQALLDKSPALRQAG
PAIR_032	FSPDNLVKVAA	GGAQALQALLDKGPALRQAG
PAIR_033	FSPDNLVKVAA	GGAHALQALLDKGPALRQAG
PAIR_034	FSPDNLVKVAA	GSQQALQALLDKSPALRQAG
PAIR_035	GGREQVIKIAA	GGQQALQALLDKGPALRNAG
PAIR_036	GGREQVIKIAA	GGQQALQALLDKGPALRQAG
PAIR_037	GGREQVIKIAA	GGKQALQALLDKSPALRQAG
PAIR_038	GSREQVIKIAA	GGQQALQALLDKGPALRNAG
PAIR_039	GSREQVIKIAA	GGQQALQALLDKSPALRQAG

[0050] TABLE 14: Pair or RULP,RURP derived from WP_22624444.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_040	TKADIVKIASN	GGMQALQAVINLHSELTKIG
PAIR_041	LSNNNIVNIAA	GGSQALRAVFTHHPALIQAG
PAIR_042	FSNDQIAKIAG	GGAQTVQAVIDLYHLLTNAG
PAIR_043	LNNKNIVKIAG	GGAQALRAVVTHHPALIEAG
PAIR_044	FSNDHVVKIGG	GGAQALQAVANLHSSLEVAG
PAIR_045	FGNNGIVRIAG	GGAQALRAVITHGSALVQRG
PAIR_046	FSNDDIVGIAG	GGAQALQAVITHYPALIQAG

[0051] TABLE 15: Pair or RULP,RURP derived from WP_13312949.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_047	FTGEQILKIVA	GGSKNLNAVLLHFKALRALK
PAIR_048	FNCKDIVKIVG	GGSKNLNAVLAHSEALCALQ
PAIR_049	FQVEDIVKILA	GGSKNLNAVLAHSEALLALQ
PAIR_050	FTGQDILKMVG	GGSKNLNAVLEHFEALRALQ
PAIR_051	FTGEDIVKIVG	GGSKNLGAVLVNFKTLRDLQ

[0052] TABLE 16: Pair or RULP,RURP derived from WP_13313355.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_052	FNPEQIVKMVS	GGSRNLDAVKNNVEALDKLG

PAIR_053	FDSNQIVKMVS	GGSKNLDAVKNNLEALDKLG
PAIR_054	FDSNQIVKMVS	GGSKNLDAVKNNAEALDKLG
PAIR_055	FDSNQIVKMVS	GGSKNLDAIKNNLEALDKLG
PAIR_056	FDSNQIVKMVS	GGSKNLDAVKNNVEILKALD
PAIR_057	FNPEQIVKMVS	GGSRNLDAVKNNVEILKALD
PAIR_058	FNPEQIVKMVS	GGSKNLDAVKNNLEILKALD
PAIR_059	FNPEQIVKMVS	GGSKNLDAVKNNLEALDKLG
PAIR_060	FDSNQIVKMVS	GGSRNLDAVKNNLEALDKLG
PAIR_061	FDSNQIVKMVS	GGSRNLDAVKNNADILKALE

[0053] TABLE 17: Pair or RULP,RURP derived from KAG0189736.1 and KAG0162681.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_062	FSKQEAVAIAS	GGSQALNTVLATHATLTAAG
PAIR_063	FTHQQIVAIAS	GGSQALNTVLATHAALTAAG
PAIR_064	FTHQQIVAIAS	GGSQALDKVLATHAPLTAAG
PAIR_065	FTHRQIVGIAS	GGSQALDTVLVRYAPLRDAG
PAIR_066	FKHEQIVGIAS	GGSQALDKVLATHAQLTAVG
PAIR_067	FKHEQIVAIAS	GGSQALDKVLVKYAPLTAAG
PAIR_068	FTHQQIVAIAS	GGSQALDTVLATHAQLTTAG
PAIR_069	FSKQEAVAIAS	GGSQALDKVLATHAQLTAVG
PAIR_070	FTHQQIVAIAS	GGSQALDKVLATHAQLTTAG

[0054] TABLE 18: Pair or RULP,RURP derived from TLY47364.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_071	YTQNQLNKIAS	GGSKTLNTLLEKAPQLLTLG
PAIR_072	YKVEQIVKVAA	GGSKTLNTLLEKTPQLLALG
PAIR_073	YKDEQLIKVAA	GGSKTLNTLLEKTPQLLALG
PAIR_074	YKDEQLIKVAA	GGSKTLNTLLEKTPQLLTLG
PAIR_075	YKDEQIVKVAA	GGSQALTTLLEKTPQLLILG

PAIR_076	YKADQLIKAAA	GGSQALNTLLEKTPQLLTLG

[0055] TABLE 19: Pair or RULP,RURP derived from TAK77069.1, and TAK78877.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_077	FTAEQMVKMVS	GGSKNLEAIKNNYDALKELG
PAIR_078	FTAEQMVKMVS	GGSKNLEAIRYGSDVLKYLG
PAIR_079	FTSEQLVDMVS	GGSKNLEELKMSYYVLKDLG
PAIR_080	FTVEQMVNMVS	GGSKNLEAIRYSSDALKYLG
PAIR_081	FTSEQMVNMVS	GGSKNLEAIRYSYHVLKELG
PAIR_082	FTTEQMVKMVK	GGSKNLEAIKNNYDALKALG
PAIR_083	FTAERMVKMAS	GGSKNLEIIKNNYDALKELG
PAIR_084	FTAEQMVKMVN	GGSRNLEAIKNNYDALKALG
PAIR_085	FTAEQMVKMAS	GGSKNLEIIKNNYDVLKESG

[0056] TABLE 20: Pair or RULP,RURP derived from WP_25856893.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_086	YSTADITRIAA	GGSKNLEAVNLKHTELISLG
PAIR_087	FNAIQIVSMVS	GGSKNLQAVTDNNEALKDLS
PAIR_088	FTAKQIVSIVS	GGSKNLQAVTENNEALKDLG
PAIR_089	FNAVQVVRMVS	GGSKNLQAVTENHEALLNLS
PAIR_090	FTAEQIVRMAS	GGSKNLQAVTENHEALLNLS
PAIR_091	FTAEQIVSMVS	GGSKNLQVVTDNNEALKDLG
PAIR_092	FNAVQVVRMVS	GGSKNLQAVTENNEALKGLG
PAIR_093	FTAVQVVRMVS	GGSKNLQAVTDNNEALKGLG
PAIR_094	FTAKQIVRMVS	GGSKNLQAITDNNEALLNLG
PAIR_095	FTAAQIVSMVS	GGSKNLQAVTDNNEALKGLG

[0057] TABLE 21: Pair or RULP,RURP derived from OGV35086.1 and OGV33042.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_096	FPREEIGKIAG	GGSHNLKAVLTHTQALINLG

PAIR_097	FPREEIGKIAG	GGSHNLEAVLTHARALIDLG
PAIR_098	FPREEIGKIVG	GGSRNLEAVLTHARALIDLG
PAIR_099	FPCNEIGKIVG	GGSRNLKAVLTHARALIDLG
PAIR_100	FLREEIGKIAG	GGSRNLKAVLTHARALIYLG
PAIR_101	FPCEEIGKIAG	GGSHNLEAVRTHVQALINLG
PAIR_102	FPREEIGKIAG	GGSHNLEAVLTHARALVDLG
PAIR_103	FPREEIGKIAG	GGSHNLEAVLTHAQALIHLG
PAIR_104	FQREEIGKIAG	GGSRNLEAVLTHAQALINLG
PAIR_105	FPCEEIGVIAG	GGSRNLDAVLTHARSLIDLG
PAIR_106	FPHEEIGKIAG	GGSRNLKAVLTHAQALIDLG
PAIR_107	FSREEISKIAG	GGSHNLEAVLKHFNVLEKLG
PAIR_108	FTHAELVKIAR	GGSRNLKAVHVNAQALIDLG
PAIR_109	FPREEVGKIAG	GGSLNLEAMLTHARALIDLG
PAIR_110	FQHEEICQIAR	GGSRNLKAVLTDAQSLIDLG
PAIR_111	FPREEISKIAG	GGSHNLAAVLKHVQTLIDLG
PAIR_112	FQREEIGKIAG	GGSLNLQAVLTNAQALIDLG
PAIR_113	FSREEIGKIAG	GGSRNLEAVNKHVQTLIDLG
PAIR_114	FQHEEISKIAG	GGSLNLQAVLTNAQALIDLG
PAIR_115	FPREDIGKIAG	GGSCNLEAMLKHFSILQKLG

[0058] TABLE 22: Pair or RULP,RURP derived from MBW8829317.1

Pair ID	Repeat Unit Left Portion (RULP)	Repeat Unit Right Portion (RURP)
PAIR_116	FSRTEIVSIAS	GGSQALGKVLATLERLKVAG
PAIR_117	FEHKHIVAIAA	GGSQALDKVLDTHERLKNAG
PAIR_118	FEHKHIVAIAS	GASQALDKVLSTHEQLKEAG
PAIR_119	FEVNQIAAIAT	GGSRALDKVLAAHKQTKAGR
PAIR_120	FDHEEIVNIAS	GGSQALAKVLATHDRLRSAG
PAIR_121	FEHEHIVAIAA	GGKQALEKVLSKHEQFKDAG

[0059] TABLE 23: Correspondence between Repeat Variable Di-residues (RVDs) and nucleotide bases.

RVD	Base
AA	T
AD	C

AG	T
AI	A
AK	G
AN	G
CD	C
CG	T
CI	A
CK	G
CN	G
CP	T
DH	G
DI	A
DN	G
ED	C
EI	A
EN	G
FI	A
FN	G
GG	T
GI	A
GN	G
H*	T
HA	C
HD	C
HG	T
HH	G
HI	A
HK	G
HN	G
HV	A, T, or G
ID	C
IG	T
IN	G
IS	T

KD	C
KG	T
KI	A
KK	G
KN	G
LN	G
MG	T
N*	C, or T
NA	G
ND	C
NG	T
NI	A
NK	G
NN	A, or G
NS	A, G, T, or C
QA	T
QG	T
QI	A
QK	G
QN	G
RD	C
RG	T
RH	G
RI	A
RN	G
SA	T
SD	C
SG	T
SI	A, or G
SN	A, or G
SW	A
TG	T
TI	A
TL	A

TN	G
VA	T
VG	T
VI	A
VN	G
VT	A, or G
WG	T
WN	G
YA	T
YG	T
YI	A
YN	G
YP	T

The * in the RVDs N* and H* denotes a gap, i.e. the residue at the second position of the RVDs is lacking.

[0060] In some embodiments, a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FGNDNLVKVAA, FGPDNLVKVAA, FGPDSLVKVAA, FSADNLVRIAA, FSLDNLVKVAA, FSNDNLIKVAA, FSNDNLMRIAA, FSNDNLVKVAA, FSPDNLIKVAA, FSPDNLVKVAA, GGREQVIKIAA, GSREQVIKIAA;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGAQALQALLDKGPALRQAG, GGAQALQALLDKGPTLRQAG, GGAQALQALLDKGTALRQAG, GGAQALQALLDRGPALRQAG, GGQHALQALLDKGPALRQAG, GGQQALQALLDKGPALRNAG, GGQQALQALLDRGPALRQAG, GSQHALQALLDKGPALRQAG, GSQQALQALLDKGPALRQAG, GSAQALQALLDKGPALRQAG, GGQQALQALLDKGPALRQAG, GGAQALQALLDKGPALLQAG, GGTQALQALLDKGPALRQAG, GGAQALQALLDKGPALRNAG, GGAQALQALLDKGSALRQAG, GGAQALQTLLDKGPALRQAG, GGAHALQALLDKGPALRQAG, GAQQALQALLDKGPALRQAG, GGQQALQTLLDKGPALRQAG, GGAQALQALLDKSPALRQAG, GSQQALQALLDKSPALRQAG, GGKQALQALLDKSPALRQAG, GGQQALQALLDKSPALRQAG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0061] In some embodiments, a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: TKADIVKIASN, LSNNNIVNIAA, FSNDQIAKIAG, LNNKNIVKIAG, FSNDHVVKIGG, FGNNGIVRIAG, FSNDDIVGIAG;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGMQALQAVINLHSELTKIG, GGSQALRAVFTHHPALIQAG, GGAQTVQAVIDLYHLLTNAG, GGAQALRAVVTHHPALIEAG, GGAQALQAVANLHSSLEVAG, GGAQALRAVITHGSALVQRG, GGAQALQAVITHYPALIQAG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0062] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FTGEQILKIVA, FNCKDIVKIVG, FQVEDIVKILA, FTGQDILKMVG, FTGEDIVKIVG;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSKNLNAVLLHFKALRALK, GGSKNLNAVLAHSEALCALQ, GGSKNLNAVLAHSEALLALQ, GGSKNLNAVLEHFEALRALQ, GGSKNLGAVLVNFKTLRDLQ;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0063] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FNPEQIVKMVS, FDSNQIVKMVS;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSRNLDAVKNNVEALDKLG, GGSKNLDAVKNNLEALDKLG, GGSKNLDAVKNNAEALDKLG, GGSKNLDAIKNNLEALDKLG, GGSKNLDAVKNNVEILKALD, GGSRNLDAVKNNVEILKALD, GGSKNLDAVKNNLEILKALD, GGSRNLDAVKNNLEALDKLG, GGSRNLDAVKNNADILKALE;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0064] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FSKQEAVAIAS, FTHQQIVAIAS, FTHRQIVGIAS, FKHEQIVGIAS, FKHEQIVAIAS;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSQALNTVLATHATLTAAG, GGSQALNTVLATHAALTAAG, GGSQALDKVLATHAPLTAAG, GGSQALDTVLVRYAPLRDAG, GGSQALDKVLATHAQLTAVG, GGSQALDKVLVKYAPLTAAG, GGSQALDTVLATHAQLTTAG, GGSQALDKVLATHAQLTTAG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0065] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: YTQNQLNKIAS, YKVEQIVKVAA, YKDEQLIKVAA, YKDEQIVKVAA, YKADQLIKAAA;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSKTLNTLLEKAPQLLTLG, GGSKTLNTLLEKTPQLLALG, GGSKTLNTLLEKTPQLLTLG, GGSQALTTLLEKTPQLLILG, GGSQALNTLLEKTPQLLTLG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0066] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FTAEQMVKMVS, FTSEQLVDMVS, FTVEQMVNMVS, FTSEQMVNMVS, FTTEQMVKMVK, FTAERMVKMAS, FTAEQMVKMVN, FTAEQMVKMAS;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSKNLEAIKNNYDALKELG, GGSKNLEAIRYGSDVLKYLG, GGSKNLEELKMSYYVLKDLG, GGSKNLEAIRYSSDALKYLG, GGSKNLEAIRYSYHVLKELG, GGSKNLEAIKNNYDALKALG, GGSKNLEIIKNNYDALKELG, GGSRNLEAIKNNYDALKALG, GGSKNLEIIKNNYDVLKESG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0067] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: YSTADITRIAA, FNAIQIVSMVS, FTAKQIVSIVS, FNAVQVVRMVS, FTAEQIVRMAS, FTAEQIVSMVS, FTAVQVVRMVS, FTAKQIVRMVS, FTAAQIVSMVS;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSKNLEAVNLKHTELISLG, GGSKNLQAVTDNNEALKDLS, GGSKNLQAVTENNEALKDLG, GGSKNLQAVTENHEALLNLS, GGSKNLQVVTDNNEALKDLG, GGSKNLQAVTENNEALKGLG, GGSKNLQAVTDNNEALKGLG, GGSKNLQAITDNNEALLNLG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0068] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FPREEIGKIAG, FPREEIGKIVG, FPCNEIGKIVG, FLREEIGKIAG, FPCEEIGKIAG, FQREEIGKIAG, FPCEEIGVIAG, FPHEEIGKIAG, FSREEISKIAG, FTHAELVKIAR, FPREEVGKIAG, FQHEEICQIAR, FPREEISKIAG, FSREEIGKIAG, FQHEEISKIAG, FPREDIGKIAG;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSHNLKAVLTHTQALINLG, GGSHNLEAVLTHARALIDLG, GGSRNLEAVLTHARALIDLG, GGSRNLKAVLTHARALIDLG, GGSRNLKAVLTHARALIYLG, GGSHNLEAVRTHVQALINLG, GGSHNLEAVLTHARALVDLG, GGSHNLEAVLTHAQALIHLG, GGSRNLEAVLTHAQALINLG,

GGSRNLDAVLTHARSLIDLG, GGSRNLKAVLTHAQALIDLG, GGSHNLEAVLKHFNVLEKLG, GGSRNLKAVHVNAQALIDLG, GGSLNLEAMLTHARALIDLG, GGSRNLKAVLTDAQSLIDLG, GGSHNLAAVLKHVQTLIDLG, GGSLNLQAVLTNAQALIDLG, GGSRNLEAVNKHVQTLIDLG, GGSCNLEAMLKHFSILQKLG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0069] In some embodiments , a repeat unit is assembled by:

(a) Selecting a RULP from any one of the sequences: FSRTEIVSIAS, FEHKHIVAIAA, FEHKHIVAIAS, FEVNQIAAIAT, FDHEEIVNIAS, FEHEHIVAIAA;

(b) Selecting one RVD from Table 23 corresponding to the nucleic acid base that the repeat unit specifically recognizes;

(c) Selecting a RURP from any one of the sequences: GGSQALGKVLATLERLKVAG, GGSQALDKVLDTHERLKNAG, GASQALDKVLSTHEQLKEAG, GGSRALDKVLAAHKQTKAGR, GGSQALAKVLATHDRLRSAG, GGKQALEKVLSKHEQFKDAG;

(d) Fusing together each part in the following order: RULP, RVD, RURP.

[0070] In some aspects, a Modular Nucleic Acid-Binding Protein (MNABP) is assembled by:

(a) Selecting SEQ ID: NTER_011, or a truncation thereof, to be the N-terminal domain;

(b) Selecting SEQ ID: RU_095 to be the first repeat unit of the Nucleic acid-Binding Domain (NBD), SEQ ID: HRU_006 to be the last repeat unit of the NBD, and one or more of SEQ ID: RU_096, SEQ ID: RU_097, SEQ ID: RU_098, SEQ ID: RU_099, SEQ ID: RU_100, SEQ ID: RU_101, SEQ ID: RU_102, SEQ ID: RU_103, or SEQ ID: RU_104 to serve as the other repeat units of the NBD. The RVD of each one repeat unit is to be substituted with the RVD set forth in Table 23;

(c) Selecting CTER_013, or a truncation thereof, to be the C-terminal domain.

(d) Fusing the parts in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain.

[0071] In some aspects, a Modular Nucleic Acid-Binding Protein (MNABP) is assembled by:

(a) Selecting SEQ ID: NTER_012, or a truncation thereof, to be the N-terminal domain;

(b) Selecting SEQ ID: RU_117 to be the first repeat unit of the Nucleic acid-Binding Domain (NBD), SEQ ID: RU_125 to be the last repeat unit of the NBD, and one or more of SEQ ID: RU_105, SEQ ID: RU_106, SEQ ID: RU_107, SEQ ID: RU_108, SEQ ID: RU_109, SEQ ID: RU_110, SEQ ID: RU_111, SEQ ID: RU_112, SEQ ID: RU_113, SEQ ID: RU_114, SEQ ID: RU_115, SEQ ID: RU_116, SEQ ID: RU_118, SEQ ID: RU_119, SEQ ID: RU_120, SEQ ID: RU_121, SEQ ID: RU_122, SEQ ID: RU_1123, or SEQ ID: RU_124 to serve as the other repeat units of the NBD. The RVD of each one repeat unit is to be substituted with the RVD set forth in Table 23;

(c) Selecting one of SEQ ID: CTER_015, or SEQ ID: CTER_016, or a truncation thereof, to be the C-terminal domain.

(d) Fusing the parts in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain

[0072] In some aspects, a Modular Nucleic Acid-Binding Protein (MNABP) is assembled by:

(a) Selecting SEQ ID: NTER_013, or a truncation thereof, to be the N-terminal domain;

(b) Selecting SEQ ID: RU_070 to be the first repeat unit of the Nucleic acid-Binding Domain (NBD), SEQ ID: HRU_008 to be the last repeat unit of the NBD, and one or more of SEQ ID: RU_071, SEQ ID: RU_072, SEQ ID: RU_073, SEQ ID: RU_074, or SEQ ID: RU_075 to serve as the other repeat units of the NBD. The RVD of each one repeat unit is to be substituted with the RVD set forth in Table 23;

(c) Selecting SEQ ID: CTER_017 or a truncation thereof, to be the C-terminal domain.

(d) Fusing the parts in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain.

[0073] In some aspects, a Modular Nucleic Acid-Binding Protein (MNABP) is assembled by:

(a) Selecting SEQ ID: NTER_014, or a truncation thereof, to be the N-terminal domain;

(b) Selecting one of SEQ ID: RU_042, SEQ ID: RU_043, or SEQ ID: RU_044 to be the first repeat unit of the Nucleic acid-Binding Domain (NBD), and one or more of SEQ ID: RU_001, SEQ ID: RU_002, SEQ ID: RU_003, SEQ ID: RU_004, SEQ ID: RU_005, SEQ ID: RU_006, SEQ ID: RU_007, SEQ ID: RU_008, SEQ ID: RU_009, SEQ ID: RU_010, SEQ ID: RU_011, SEQ ID: RU_012, SEQ ID: RU_013, SEQ ID: RU_014, SEQ ID: RU_015, SEQ ID: RU_016, SEQ ID: RU_017, SEQ ID: RU_018, SEQ ID: RU_019, SEQ ID: RU_020, SEQ ID: RU_021, SEQ ID: RU_022, SEQ ID: RU_023, SEQ ID: RU_024, SEQ ID: RU_025, SEQ ID: RU_026, SEQ ID: RU_027, SEQ ID: RU_028, SEQ ID: RU_029, SEQ ID: RU_030, SEQ ID: RU_031, SEQ ID: RU_032, SEQ ID: RU_033, SEQ ID: RU_034, SEQ ID: RU_035, SEQ ID: RU_036, SEQ ID: RU_037, SEQ ID: RU_038, SEQ ID: RU_039, SEQ ID: RU_040, SEQ ID: RU_041, SEQ ID: RU_045, or SEQ ID: RU_046 to serve as the other repeat units of the NBD. The RVD of each one repeat unit is to be substituted with the RVD set forth in Table 23;

(c) Selecting one of SEQ ID: CTER_018, SEQ ID: CTER_019, or SEQ ID: CTER_020 or a truncation thereof, to be the C-terminal domain.

(d) Fusing the parts in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain.

[0074] In some aspects, a Modular Nucleic Acid-Binding Protein (MNABP) is assembled by:

(a) Selecting SEQ ID: NTER_002, or a truncation thereof, to be the N-terminal domain;

(b) Selecting SEQ ID: RU_059 to be the first repeat unit of the Nucleic acid-Binding Domain (NBD), SEQ ID: HRU_014 to be the last repeat unit of the NBD, and one or more of SEQ ID: RU_060, SEQ ID: RU_061,

SEQ ID: RU_062, SEQ ID: RU_063, SEQ ID: RU_064, SEQ ID: RU_065, SEQ ID: RU_066, SEQ ID: RU_067, SEQ ID: RU_068, or SEQ ID: RU_069 to serve as the other repeat units of the NBD. The RVD of each one repeat unit is to be substituted with the RVD set forth in Table 23;

(c) Selecting SEQ ID: CTER_003 or a truncation thereof, to be the C-terminal domain.

(d) Fusing the parts in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain.

[0075] In some aspects, the Modular Nucleic Acid-Binding Protein (MNABP) comprises a Nucleic acid-Binding Domain (NBD) comprising at least four independent repeat units ordered from the N-terminus to the C-terminus of the said NBD, each of said repeat units having specificity for a nucleotide base of a target sequence.

[0076] In some aspects, a Modular Nucleic Acid-Binding Proteins (MNABP) comprises a Nucleic acid-Binding Domain (NBD) comprising, ordered from N-terminus to C-terminus, at least three independent repeat units and an independent half-repeat unit, each of said repeat unit having specificity for a nucleotide base of a target sequence, and the said half-repeat units having specificity for the last nucleotide base of the said target sequence.

[0077] In certain aspects, the half-repeat unit of the Nucleic acid-Binding Domain (NBD) is selected from one of the sequences provided in Table 24. The RVD of the selected half-repeat unit (HRU) can be substituted with an RVD set forth in Table 23.

[0078] TABLE 24: Exemplary half-repeat units.

SEQ ID	Half-repeat unit sequence	RVD	Derived from
HRU_001	FTAEQMVKIFS**HN**GGSRTLEVLLNRINIFDFIG	HN	TAK78877.1
HRU_002	FSREEISKIAG**HG**GGSHNLEAVLKHFNVLEKLG	HG	OGV35086.1
HRU_003	FPREDIGKIAG**RD**GGSCNLEAMLKHFSILQKLG	RD	OGV33042.1
HRU_004	FNAEQIVRMVS**HK**GGSKNLALVKEYFPVFSSFH	HK	WP_058473422.1
HRU_005	FEIEDIVAMAS**HV**GGAPAMQSILDHLDILQAHY	HV	MBW8829317.1
HRU_006	FNTIQIVSMVS**HD**GGSKNLQAVTAGYEKLSKVW	HD	WP_258568932.1
HRU_007	LTPAQVVAIAS**NG**GGKQALESIVAQLSRPDPALA	NG	WP_263108675.1
HRU_008	FAVEDVSAIAA**HI**GGAPALQAVVDHLELLMTRH	HI	KAG0162681.1
HRU_009	FAVEDVSAIAA**HI**GGAPALQAVVDHLELLMTRH	HI	KAG0189736.1
HRU_010	FSAADIVKIAS**NN**GGAQALQALIDHWSTLSGKT	NN	OXJ06552.1
HRU_011	FSAADIVKIAS**NN**GGARALQALIDHWSTLSGKT	NN	OXJ06553.1
HRU_012	FSAADIVKIAS**NN**GGARALQALIDHWSTLSGKT	NN	WP_089477027.1
HRU_013	FSAEQIVRIAA**HI**GGSRNIEATIKHYAMLTQPP	HI	WP_058451450.1
HRU_014	FDSNQIVKMVS**HI**GGSKNLDSVLKLAELIDDND	HI	WP_133133554.1
HRU_015	LTPNQVVAIAS**NG**GGKQALESIVAQLSRPDPAL	NG	AIA22682.1

HRU_016	FGNDNLVRIGG**NG**GAKKTLDTLLQVYPQLTQGG	NG	WP_168083840.1
HRU_017	FSNDNLVRIGG**NG**GAKKTLDTLLQVYPKLTQGG	NG	WP_178089108.1

[0079] The repeat units comprising RVDs are fused together from the N-terminus to the C-terminus according to the sequence of the nucleic acid that we desire to target. The ordering of the RVDs of the assembled repeat units is called the "RVD sequence".

[0080] In some aspects, the Nucleic acid-Binding Domain (NBD) of the herein Modular Nucleic Acid-Binding Protein (MNABP) comprises at least three herein disclosed Repeat Units, and at least one Repeat Unit (RU) selected from any one of the Repeat Units disclosed in patent WO2019204643A2 (Urnov F et al., 2018), wherein the RVD of the therein selected Repeat unit is substituted with an RVD set forth in Table 23.

[0081] In some aspects, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises an N-terminal domain, wherein the C-terminus of the N-terminal domain is fused to the N-terminus of the first repeat unit of the nucleic acid binding domain (NBD) of the MNABP.

[0082] In certain aspects, the N-terminal domain of the MNABP can be SEQ ID: NTER_001 (see Table 25). In a particular embodiment, the sequence of the first repeat unit of the nucleic acid binding domain (NBD) can be FSSQQIIRMVS**XX**GGANNLKAVTANHDDLQNMG, wherein the **XX** is substituted with an RVD (provided herein in Table 23) in respect of the nucleotide base to which the said first repeat unit binds.

[0083] I some aspects, the sequence of the N-terminal domain of the MNABP can be SEQ ID: NTER_002 (see Table 25). In a particular embodiment, the sequence of the first repeat unit of the nucleic acid binding domain (NBD) can be FNPEQIVKMVS**HD**GGSRNLDAVKNNVEALDKLG, wherein the **XX** is substituted with an RVD (provided herein in Table 23) in respect of the nucleotide base to which the said first repeat unit binds.

[0084] In certain aspects, the MNAPB comprises an N-terminal domain that was generated by following the teaching set forth in Patent US9902962B2 (Barbas et al., 2012), wherein the C-terminus of the therein disclosed N-terminal domain is fused to the N-terminus of the herein disclosed nucleic acid binding domain (NBD), and wherein the first repeat unit of the NBD mediates the specific recognition of the second nucleotide base of the target nucleic acid to which the MNAPB binds.

[0085] In certain aspects, the MNAPB comprises an N-terminal domain that was generated by following the teaching set forth in Patent EP2780460B1 (Gregory et al., 2011), wherein the C-terminus of the therein disclosed N-terminal domain is fused to the N-terminus of the herein disclosed nucleic acid binding domain (NBD), and wherein the first repeat unit of the NBD mediates the specific recognition of the second nucleotide base of the target nucleic acid to which the MNAPB binds.

[0086] Table 25: Exemplary N-terminal domains.

SEQ ID	N-terminal domain	Derived from
NTER_001	MPDLELNFAIPLHLFDDETVFTHDATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSKE PANTKSRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSAVQQL LCKHEELLNLISPDGLGHKELIKIAARNGGGNNLIAVLSCYAKLKEMG	WP_058473422.1
NTER_002	MTNTTSKKSKYSLENLAKYNIKVEEYDKAKKELSTRGYEEYQIEKIVIRKSYRRSYLKLLELHEILV DEIKLTHDQITNIARKGGGSKNLESIKNNFKLLKTLK	WP_133133554.1
NTER_003	MPKTNQPKNLEAKSTKNKISLPQDPQTLNELKIKGYPQDLAERLIKKGSSLAVKTVLKDHEQLV NFFTHLQIIRMAAQKGGAKNITTALNEYNSLTNLG	WP_058451450.1
NTER_004	MPATSMHQEDKQSANGLNLSPLERIKIEKHYGGGATLAFISNQHDELAQVLSRADILKIASYD CAAQALQAVLDCGPMLGKRG	WP_013436752.1
NTER_005	MSTAFVDQDKQMANRLNLSPLERSKIEKQYGGATTLAFISNKQNELAQILSRADILKIASYDCA AHALQAVLDCGPMLGKRG	WP_013428821.1
NTER_006	MPVTSVYQKDKPFGARLNLSPFECLKIEKHSGGADALEFISNKYDALTQVLSRADILKIACHDC AAHALQAVLDYEQVFRQRG	WP_013436750.1
NTER_007	MSAMFMPQEGKQSANGLNLSPLERIKIEKHYGGSATLAFISNQHDELAQVLSRADILKIASYD CAAQALQAVLDCGPMLGKRG	SIT71710.1
NTER_008	MPATFMHQEDKQSANGLNLSPLERSKIEKHYGGAATLAFISNQHDELAQVLSRTDILKIASYD CAAQALQAVLDCGPMLGKRG	SIT73265.1
NTER_009	MPVTSVYQKDKPFGARLNLSPLERIKIEKHYGGSATLEFISNQHDKLAQVLSRADILKIASYDCA AQALQAVLDCGPMLGKRG	SIT64975.1
NTER_010	MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSP AFSAGSFSDPLRQFDPSLLDTSLFDSMPAVGTPHTAAAPAEWDEAQSALRAADDPPPTVRVA VTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVG HGFTHAHIVALSKHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSGARALEALLTKAGEL RGPPLQLDTGQLLKIAKRGGVTAVEAVHASRNALTGAPLN	WP_263108675.1
NTER_011	MKRTAEHRQETIKKLKLDTTWVGLNKSILDLGFTQAQADKIILRRSSGNTITAVLKHTTRLVSLG	WP_258568932.1
NTER_012	MRRKTLIDVNGLTAHLEKFNISKARYTKEVGDLRAKGYLEIEAQTVIFRKGSEKTVETLLDLHDA LIAQE	OGV33042.1
NTER_013	MDIRSLLNPLPSPGPGERAPGKRASDATPRALPSSLPDFGLPQGKRRKTTVGSSPGGRPRQDL STLSAFFQRARVSEDAHPASATVEQSGPLGATNWILSGQETNRIKKSGGAKALETLSEKAEAL HRAG	KAG0189736.1
NTER_014	MNEWIQRHNPQDKQQSSGASVSTQSMVFSQAGSANVSAGVPGPSRTRATHTDTHTVRHS PYPAASARSATSARSANTSSQALSTADHKKIQKAAGNATLNYVIQHLDELQHAL	WP_168083840.1

Derived from: The values are protein NCBI accession numbers.

[0087] In certain aspects, the N-terminal domain of the MNABP is selected from one of SEQ ID: NTER_003, SEQ ID: NTER_004, SEQ ID: NTER_005, SEQ ID: NTER_006, SEQ ID: NTER_007, SEQ ID: NTER_008, SEQ ID: NTER_009, SEQ ID: NTER_010, SEQ ID: NTER_011, SEQ ID: NTER_012, SEQ ID: NTER_013, or SEQ ID: NTER_014 (see Table 25).

[0088] In some aspects, the N-terminal domain is preferably a truncated N-terminal domain. A truncated N-terminal domain can be obtained by deleting at least one residue from the N-terminus of any one of the sequences provided herein in Table 25, wherein the resulting truncated N-terminal domain has a length of 49 residues or more. For example, a truncated N-terminal domain can be obtained by deleting the firsts 14 residues from the N-terminus of SEQ ID: NTER_001 (sequence: **MPDLELNFAIPLHLFDDETVFTH**DATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSKEPANTKSRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSAVQQLLCKHEELLNLISPDGLGHKELIKIAARNGGGNNLIAVLSCYAKLKEMG. Residues in bold and underlined are the first 14 residues of the sequence), giving a truncated N-terminal domain with a length of 153 residues and a sequence of DATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSKEPANTKSRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSAVQQLLCKHEELLNLISPDGLGHKELIKIAARNGGGNNLIAVLSCYAKLKEMG.

[0089] In some aspects, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises a C-terminal domain, wherein the N-terminus of the C-terminal domain is fused to the C-terminus of the last repeat unit or the half-repeat unit of the nucleic acid binding domain (NBD) of the MNABP.

[0090] In some aspects, the C-terminal domain of the Modular Nucleic Acid-Binding Protein (MNABP) is selected from one of the sequences provided herein in Table 26.

[0091] In certain aspects, the C-terminal domain is a C-terminal domain derived from Ralstonia protein CAD15517.1 (Table 26, SEQ ID: CTER_010).

[0092] In certain aspects, the C-terminal domain is a “C-terminal domain of an endogenous TALE molecule” disclosed in patent US9499592B2 (Zhang et al., 2011), or a truncation thereof.

[0093] In some aspects, the C-terminal domain is preferably a truncated C-terminal domain. A truncated C-terminal domain can be obtained by deleting at least one residue from the C-terminus of any one of the sequences provided herein in Table 26, wherein the resulting truncated C-terminal domain has a length of 10 residues or more. For example, a truncated C-terminal domain can be obtained by deleting the last 267 residues from the C-terminus of SEQ ID: CTER_010 (sequence: LSPERVAAIACIGGRSAVEAVRQGLPVKAI**RRIRREKAPVAGPPPASLGPTPQELVAVLHFFRAHQQPRQAFVDALAAFQATRPALLRLLSSVGVTEIEALGGTIPDATERWQRLLGRLGFRPATGAAAPSPDSLQGFAQSLERTLGSPGMAGQSACSPHRKRPAETAIAPRSIRRSPNN**

AGQPSEPWPDQLAWLQRRKRTARSHIRADSAASVPANLHLGTRAQFTPDRLRAEPGPIMQAHTSPASVSFGSHVAFEPGLPDPGTPT SADLASFEAEPFGVGPLDFHLDWLLQILET. Residues in bold and underlined are the last 267 residues of the sequence), giving a truncated C-terminal domain with a length of 30 residues and a sequence of LSPERVAAIACIGGRSAVEAVRQGLPVKAI.

[0094] Table 26: Exemplary C-terminal domains.

SEQ ID	C-terminal domain	Derived from
CTER_001	YMLSQEQFLRLIDHHSGHLNLSILLDEQQWQAINDLCLQPHHFGRQNALEKFLQQGQRK YQNLQELEQFLFQDSADPMLLQETENQHEAEKINDCMDFILRLISATEPLDLQIEIEGIGLFS PSMHFDATQANFSTPAANEEKIDNSATEAGVNSRKRKIAAAHQKQPPRKKTATPLSATFI STLTTLAQSDNPRLEMASAEALMLKAPQKLAMGITVRKKTKCEGIAIITVTDKTKLNGWLS SASESTYSSVEAQGTRTVNNTHAFFSTPLTSDKKSPSFSSLDFYEDSGLGFDEEITNPPYMP ELEPEFIL	WP_058451450.1
CTER_002	FTADQIVALICQSKQCFRNLKKNHQQWKNKGLSAEQIVDLILQETPPKPNFNNTSSSTPSP SAPSFFQGPSTPIPTPVLDNSPAPIFSNPVCFFSSRSENNTEQYLQDSTLDLDSQLGDPTKN FNVNNFWSLFPFDDVGYHPHSNDVGYHLHSDEESPFFDF	WP_058473422.1
CTER_003	LIVSEFSNKQGRKKLYDLATNKLMTSLNVTDLHLPDQIRIQVSDLTFLLEDLEIVQESIVIDPE LEVEVEVEVEVEVEVEVEVEAEHAKRTMDVTQPQNKNKRRKVQTEKQMTATIELALE VDNRHQTSHDYTSYPFTNASELFEFQDEGNIDKSVSSLDTRLNQTVNNQERNSSLPPAIEV FTYSPQNLIANTNSFSEAANTGNTQQSFLDETENDVIYDFVNSFPRNVAFDGDDNNQDLY TTDLVADANEVDNNQSNIVKDSRSSRNKLNQEAYPDYISLCCQENLESSPYNLSQIINESNII ELSDEILSELQISSQQYLSTEDTVNTREKNQMAFEIDDEKVQTRFNFMEFSSKNAESQYQQ EITDLSLENDSEYGHFNPQLDEYSILESQLFQPFFGNEVHDPAPYTTSKPNPNTSFIKETSLD TRNNFWSNCNNFFQSRQESVSNDENLIEYTNTKP	WP_133133554.1
CTER_004	RSNEEIVHVAARRGGAGRIRKMVASLLGGNRDGVTSIEGQ	SIT73265.1
CTER_005	RSNEEIVHVAARRGGAGRIRKMVAPLLGGNRDGVTSIEGP	SIT71710.1
CTER_006	RSNEEIVNVAARRGGAGRIRKMVAPLLGRQ	SIT64975.1
CTER_007	RSNEDIVNMAARTGAAGQIRKMAAQLSGRQ	WP_013436750.1
CTER_008	RSNEEIVHVAARRGGAGRIRKMVALLLERQ	WP_013436752.1
CTER_009	RSNEEIVHVAARRGGAGRIRKMVAPLLERQ	WP_013428821.1
CTER_010	LSPERVAAIACIGGRSAVEAVRQGLPVKAIRRIRREKAPVAGPPPASLGPTPQELVAVLHFF RAHQQPRQAFVDALAAFQATRPALLRLLSSVGVTEIEALGGTIPDATERWQRLLGRLGFR PATGAAAPSPDSLQGFAQSLERTLGSPGMAGQSACSPHRKRPAETAIAPRSIRRSPNNAG	CAD15517.1

	QPSEPWPDQLAWLQRRKRTARSHIRADSAASVPANLHLGTRAQFTPDRLRAEPGPIMQ AHTSPASVSFGSHVAFEPGLPDPGTPTSADLASFEAEPFGVGPLDFHLDWLLQILET	
CTER_011	LSTGQAVALACIGGRPALETARHTQAPIRRMQISPASNPAPTPTRYGPTPAQCVEVLQFF HDYLPPRSSAFADAKAKFQVSRVDLLRLLASLGVTEAEALSGTLPDAGLRWQRLLNRLNLP PRADAAQPSSAGAMQGFAESLERSLESPSPVRNSALAHHAASTGPENAEGFDLGGSGTL EELSAAQLIAGFKQTEVAFDQ	WP_193727545.1
CTER_012	EDKISRDQIADFLHKGNSGRKELYDLMISLLEKDSQDDFQEALTPTGDMIDNPTDDEDNN NNCHKRKCNSINKNEKKKRTKINPQEAVATAISLLSKFDPKQLLNDPSFILDSESTTCLKSLP KKLRDEIGIKSIRTNNKKSINTFKVMVDLNSYQPWYEIHILPSEIEMNMDDESIHSDDMLD ENIIEEAYQSIINNPIDDSCGYYKQDNPLLFFYREASSAAQLVVAEQNKDIANK	TAK78877.1
CTER_013	SHVLISSLVSQNGGSKRIYQQLKSLDAAHTLVALWPEIDADIIGDLEFQSSSSSLGLN	WP_258568932.1
CTER_014	ALTNDHLVALACLGGRPALDAVKKGLPHAPELIRRVNSRIGERTSHRVADYAQVVRVLEFF QCHSHPAYAFDEAMTQFGMSRNGLLQLFRRVGVTELEARGGTLPPASQRWDRILQASG MKRAKPSPTSAQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSRSDRAVTGP SAQQSFEVRVPEQRDALHLPLSWRVKRPRTRIGGGLPDPGTPMAADLAASSTVMWEQD AAPFAGAADDFPAFNEEELAWLMELLPQSGSVGGTI	WP_263108675.1
CTER_015	LTPENMIKYLLQPKRAAATVDLECALPIVRTTQKRERMSHDALSLGSCLPAMDEDVEVLDI SLLSEESFDAWFTSIDDDYFVFDDIGETHNMQSWEKLDELPRHGSQASSDTRTSSCVTTSF ASSSKSFWAPKNNSTVADNDVVHDNKRHKMLMG	OGV33042.1
CTER_016	LTPENMIKYLLQPKRAAATVDLECALPIVRTTQKRERMSHDALSLGSCLPAMDEDVEVLDI SLLSEESFDAWFTSIDDDYFVFDDIGETHNMQSWEKLDELPRHGSQASSDTRTSSCVTTSF ASSSKSFWAPKNNSTVADNDVVHDNKRHKMLMG	OGV33042.1
CTER_017	SKEDIVKAGAKQRGAAAHVKQMANACRIKQESAAQSPRPMPTVLVERPIDQARTAFIPEL QHCDLTGGTPIWSLDEASRVVLRHPMDPIEGNNDLFPLRDLTRPLDRVYERYADKNGKC HPNVKLTNIDLASGYKKYFNELCRDSRVGLSPSETANVRGRLLTNARTEFERLIREEAAPER PCKVRQLDHGGLLEHERMLAGQYGLFLAPAHSPQDQCTLRNGRILGFYMGMFAANEQ QINAIEAQHPDYESYAMDAMRPGGKLTVYSALGCANDLAFANTALCADTPEPAYDRERL NAEFIPFEVKLTDRHGKPARETVVAMVALDNAIGKEIRVDYGDAFLRQFTTPRDRARSEED AVVVKMEVDD	KAG0189736.1
CTER_018	VSHDEILALATKQRGASGALQSKLGELTAAGR	WP_168083840.1
CTER_019	VSQAEILTLATKHRGASGTLQSRLKELTATGR	WP_229632017.1
CTER_020	VSQAEILTLATKHRGASGTLQSKLKELTRLLGGFA	WP_195755912.1
Derived from: The values are protein NCBI accession numbers		

[0095] In some embodiments, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises at least one nuclear localization signal (NLS) located at its N-terminus, or at its C-terminus, or at both at its C- and N-terminus. In some aspects, the sequence of the nuclear localization signal can be PKKKRKV.

[0096] In certain aspects, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises a signal peptide at its N-terminus to translocate the MNABP to a specific cell compartment, wherein the signal peptide is cleaved during translocation.

[0097] In various aspects, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises a functional domain. In a particular embodiment, the N-terminus of the functional domain is fused to the C-terminus of the C-terminal domain of the MNAPB. In another embodiment, the C-terminus of the functional domain is fused to the N-terminus of the N-terminal domain of the MNAPB.

[0098] In certain aspects, the Modular Nucleic Acid-Binding Protein (MNABP) of the present invention comprises two functional domain, where the N-terminus of one of the functional domains is fused to the C-terminus of the C-terminal domain of the MNAPB, and the C-terminus of the other functional domain is fused to the N-terminus of the N-terminal domain of the MNAPB.

[0099] Non-limiting examples of functional domains are: transcriptional activation domains, transcriptional repression domains, transcriptional co-activator domains, transcriptional co-repressor domains, chromatin-modifying domains, DNA modifying domains, transposase domains, or nuclease domains.

[0100] Example of transcriptional activation domains are: AP-2, AtHD2A, CBP, CTF1, ERF-2, Oct 1, Oct-2A, p300, p65, PCAF, Sp1, SRC1 PvALF, VP16, and VP64.

[0101] Example of transcriptional repression domains are: Rb, KOX, SID, KRAB, LSD1, MBD2, MBD3, DNMT1, MeCP2, Sin3a, v-erbA, DNMT3B, SUV39H1, G9A (EHMT2), DNMT3A-DNMT3L, ROM2, AtHD2A, and TGF-beta-inducible early gene (TIEG).

[0102] Example of nuclease domains are: endonuclease domain of FokI, I-AniI, I-OnuI, or BfiI.

[0103] Example of chromatin-modifying domains are: lysine-specific histone demethylase 1, or any polypeptides with kinase, acetylase, or deacetylase activity.

[0104] Example of DNA modifying domains are: polypeptides with methyltransferase, topoisomerase, helicase, ligase, kinase, phosphatase, polymerase, or endonuclease activity.

[0105] Example of transposase domains are: Sleeping beauty transposases, PiggyBac transposases, frog prince transposases, Tol2 transposases.

[0106] In a particular embodiment, the MNABP can be fused to a DNA-endonuclease from the FokI polypeptide or a variant thereof to generate Modular Nucleic Acid-Binding ENdonucleases (MNABENs).

[0107] In a particular embodiment, the MNABP can be fused to a transposase polypeptide or a variant thereof, to generate Modular Nucleic Acid-Binding TRansposase (MNABTRs).

[0108] In some embodiments, a Modular Nucleic Acid-Binding Protein (MNABP) and a functional domain may be linked together using any suitable peptide linker sequences. Examples of peptide linker sequences are provided in Table 27.

[0109] In some aspects, the peptide linker sequence comprise a protease-cleavable domain.

[0110] TABLE 27: Exemplary peptide linkers.

SEQ ID	Peptide linker, sequence
LNKR_001	SG
LNKR_002	NVG
LNKR_003	DSVI
LNKR_004	IVEA
LNKR_005	LEGS
LNKR_006	YTST
LNKR_007	LQENL
LNKR_008	VGRQP
LNKR_009	LGNSL
LNKR_010	QGPSG
LNKR_011	LPEEKG
LNKR_012	QTYQPA
LNKR_013	FSHSTT
LNKR_014	GYTYINP
LNKR_015	LTKYKSS
LNKR_016	GRSGSDP
LNKR_017	SRPSESEG
LNKR_018	PELKQKSS
LNKR_019	LTTNLTAF
LNKR_020	LGPDGRKA
LNKR_021	LDNFINRPV
LNKR_022	VSSAKTTAP
LNKR_023	TATPPGSVT
LNKR_024	SITKSKISGS
LNKR_025	DSKAPNASNL
LNKR_026	KRRTTISIAA
LNKR_027	APAETKAEPT
LNKR_028	PVKMFDRHSSL

LNKR_029	YTRLPERSELPAEI
LNKR_030	VSTDSTPVTNQKSS
LNKR_031	YKLPAVTTMKVRPA
LNKR_032	IARTDLKKNRDYPLA
LNKR_033	SGGSGSNVGSGSGSG
LNKR_034	GGGGGMDAKSLTAWS
LNKR_035	SGGSGSDSVISGSGSG
LNKR_036	SGGSGSLEGSSGSGSG
LNKR_037	SGGSGSIVEASGSGSG
LNKR_038	SGGSGSYTSTSGSGSG
LNKR_039	SGGSGSLQENLSGSGSG
LNKR_040	SGGSGSVGRQPSGSGSG
LNKR_041	SGGSGSLGNSLSGSGSG
LNKR_042	SGGSGSQTYQPASGSGSG
LNKR_043	SGGSGSLPEEKGSGSGSG
LNKR_044	SGGSGSFSHSTTSGSGSG
LNKR_045	SGGSGSGYTYINPSGSGSG
LNKR_046	SGGSGSLTKYKSSSGSGSG
LNKR_047	SGGSGSLTTNLTAFSGSGSG
LNKR_048	GSDITKSKISEKMKGGGPSG
LNKR_049	SGGSGSSRPSESEGSGSGSG
LNKR_050	SGGSGSPELKQKSSSGSGSG
LNKR_051	TEEPGAPLTTPPTLHGNQARA
LNKR_052	SGGSGSTATPPGSVTSGSGSG
LNKR_053	ARFTLAVGDNRVLDMASTYFD
LNKR_054	SGGSGSVSSAKTTAPSGSGSG
LNKR_055	SGGSGSLDNFINRPVSGSGSG
LNKR_056	SGGSGSDSKAPNASNLSGSGSG
LNKR_057	SGGSGSKRRTTISIAASGSGSG
LNKR_058	SGGSGSPVKMFDRHSSLSGSGSG
LNKR_059	SGGSGSAPAETKAEPMTSGSGSG
LNKR_060	GSDITKSKISEKMKGLGPDGRKA
LNKR_061	IWLNRAETPLPLDPTGKVKAELDTR
LNKR_062	SGGSGSYTRLPERSELPAEISGSGSG

LNKR_063	SGGSGSVSTDSTPVTNQKSSSGSGSG
LNKR_064	SGGSGSYKLPAVTTMKVRPASGSGSG
LNKR_065	ELAEFHARYADLLLRDLRERSGSGSG
LNKR_066	DIFDYYAGVAEVMLGHIAGRSGSGSG
LNKR_067	SGGSGSIARTDLKKNRDYPLASGSGSG
LNKR_068	AAGASSVSASGHIAPLSLPSSPPSVGS
LNKR_069	ILNKEKKAVSPLLLTTTNSSEGLSMGNY
LNKR_070	ELAEFHARYADLLLRDLRERPVSLVRGPDSG
LNKR_071	ELAEFHARPDPLLLRDLRERPVSLVRGLGSG
LNKR_072	DIFDYYAGVAEVMLGHIAGRPATRKRWPNSG
LNKR_073	DIFDYYAGPDPVMLGHIAGRPATRKRWLGSG
LNKR_074	AAGGSALTAGALSLTAGALSLTAGALSGGGGS
LNKR_075	SGGSGSARFTLAVGDNRVLDMASTYFDSGSGSG
LNKR_076	SGGSGSTEEPGAPLTTPPTLHGNQARASGSGSG
LNKR_077	VAQLSRPDPAAVSAQKAKAACLGGRPALDAVKKGL
LNKR_078	SIVAQLSRPDPALVSFQKLKLACLGGRPALDAVKKGL
LNKR_079	SIVAQLSRPDPAVVTFHKLKLACLGGRPALDAVKKGL
LNKR_080	SGGSGSIVVLNRAETPLPLDPTGKVKAELDTRSGSGSG
LNKR_081	SIVAQLSRPDPAIHKKFSSIQMACLGGRPALDAVKKGL
LNKR_082	SGGSGSILNKEKKAVSPLLLTTTNSSEGLSMGNYSGSGSG
LNKR_083	SIVAQLSRPDPALQLPPLERLTLDACLGGRPALDAVKKGL
LNKR_084	SIVAQLSRPDPAAAAATNDHAVAAACLGGRPALDAVKKGL
LNKR_085	SIVAQLSRPDPAQSLAQELSLNESQIKIACLGGRPALDAVKKGL

METHODS OF PRODUCTION, DELIVERY, AND USES OF THE POLYPEPTIDES DISCLOSED HEREIN

[0111] The polypeptides disclosed herein can be produced using methods of polypeptides' productions that are well known in the prior art.

[0112] The repeat units, the half-repeat units, the N-terminal domain, the C-terminal domain, the Nucleic acid-binding Domain (NBD), the peptide linkers, the Modular Nucleic Acid-Binding Proteins (MNABPs), and any functional domains (made of polypeptides) fused to one or more MNABPs at their C- and/or N-terminus, or fragments thereof, are polypeptides that can be encoded in DNA or RNA sequences. Fusing a plurality of DNA (or RNA) together to create DNA (or RNA) sequences that encode the polypeptides disclosed herein is a practice well known in the prior art.

[0113] The novel Modular Nucleic Acid-Binding Proteins (MNABPs) disclosed herein are polypeptides that bind to a desired nucleic acid sequence. They can be used and/or delivered similarly to other well-known nucleic acid-binding polypeptides. Non-limiting examples of disclosures that teach the uses and/or delivery of nucleic acid-binding proteins are: patent US9017967B2 (Bonas et al. 2009), patent WO2011072246A2 (Voytas et al., 2010), patent US9499592B2 (Zhang et al., 2011), patent US9902962B2 (Barbas et al., 2012), patent WO2014167058A1 (Ralf et al., 2014), patent WO2019204643A2 (Urnov F et al., 2018).

[0114] The repeat units (RUs) of the present invention can be used to construct: (a) the so-called "modular base-per-base specific nucleic acid binding domains (MBBBD)" disclosed in patent WO2014018601A2 (Bertonati et al., 2012); or (b) the so-called "modular nucleic acid binding domain derived from an animal pathogen protein (MAP-NBD)" disclosed in patent WO2019204643A2 (Urnov et al., 2018).

[0115] To facilitate the construction of plasmids that encode a Modular Nucleic Acid-Binding Endonuclease fused to a functional domain, we can construct a DNA polynucleotide that comprises, ordered from 5' to 3': (a) a nuclear localization signal (NLS); (b) a restriction site (RS_FD1); (c) a sequence encoding a N-terminal domain - or a truncation thereof - selected among any one of the sequences disclosed in Table 25; (d) a restriction site (RS_RP); (e) a sequence encoding a C-terminal domain - or a truncation thereof - selected among any one of the sequences disclosed in Table 26; and (f) a restriction site (RS_FD2). The herein DNA polynucleotide can be cloned into a plasmid vector, and its expression can be driven by a strong and constitutive promoter (ex: CMV promoter). RS_FD1 and/or RS_FD2 will allow for the insertion of DNA sequences encoding a functional domain. RS_RP will allow for the insertion of a DNA sequence encoding a Nucleic acid-Binding Domain (NBD). The sequence encoding the N-terminal domain, the C-terminal domain, and the nucleic acid-binding domain should not comprise RS_RP, RS_FD1, and RS_FD2 restriction sites.

EXAMPLES

Example 1

[0116] We seek to assemble a plurality of repeat units (i.e. a Nucleic acid-Binding Domain or NBD) that specifically recognize the nucleotide sequence 5'-CGTA-3'. We select from Table 12, the RVDs HD, NK, NG, and NI, for the recognition of the bases C, G, T, and A, respectively. Each base of the target nucleotide sequence is replaced by their corresponding RVD, giving a RVD sequence of HD-NK-NG-NI. We select four repeat units from Table 1, one for each one RVD of the RVD sequence, yielding FGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAG (SEQ ID: RU_001), FGPDNLVKVAA**NK**GGQQALQALLDKGPALRQAG (SEQ ID: RU_017), FGNDNLVKVAA**NG**GGAQALQALLDKGPALRQAG (SEQ ID: RU_005), and FGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAG (SEQ ID: RU_006). We fuse the repeat units together, yielding

FGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGPDNLVKVAA**NK**GGQQALQALLDKGPALRQAGFGNDNLVKVAA**NG**GGAQALQALL
DKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAG

Example 2

[0117] We seek to assemble a repeat unit that specifically recognizes the base T. We select the pair PAIR _001 (RULP: FGNDNLVKVAA, RURP: GGAQALQALLDKGPALRQAG) from Table 13. We select **HG** from Table 12 to serve as the RVD. We fuse together each part in the following order: RULP, RVD, RURP. The resulting sequence of the repeat unit that specifically recognizes the base T is: FGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAG

Example 3

[0118] We seek to assemble a repeat unit that specifically recognizes base A. We select the pair PAIR _064 (RULP: FTHQQIVAIAS, RURP: GGSQALDKVLATHAPLTAAG) from Table 17. We select **NI** from Table 12 to serve as the RVD. We fuse together each part in the following order: RULP, RVD, RURP. The resulting sequence of the repeat unit that specifically recognizes base A is: FTHQQIVAIAS**NI**GGSQALDKVLATHAPLTAAG

Example 4

[0119] We seek to assemble a repeat unit that specifically recognizes the base G. We select FGPDSLVKVAA as the RULP, we select **HN** as the VRD according to Table 23, and we select GGAQALQALLDKGPTLRQAG as the RURP. We fuse together each part in the following order: RULP, RVD, RURP. The resulting sequence of the repeat unit that specifically recognizes the base G is: FGPDSLVKVAA**HN**GGAQALQALLDKGPTLRQAG

Example 5

[0120] We seek to assemble a repeat unit that specifically recognizes the base G. We select FSNDNLVKVAA as the RULP, we select **HK** as the VRD according to Table 23, and we select GGQQALQALLDKGPALRQAG as the RURP. We fuse together each part in the following order: RULP, RVD, RURP. The resulting sequence of the repeat unit that specifically recognizes the base G is: FSNDNLVKVAA**HK**GGQQALQALLDKGPALRQAG

Example 6

[0121] We desire to target a nucleic acid sequence 5'-AATGTACGTTA-3' of length 11 (this sequence comprises 11 nucleotide bases). The Repeat Variable Di-residue (RVD) that corresponds to a given nucleotide base is selected according to Table 23. We chose the Repeat Variable Di-residues (RVDs) NI, HG, HD, and HK, to target the nucleotide bases A, T, C, and G, respectively. We replace each nucleotide base of the sequence AATGTACGTTA by its corresponding RVD, producing the following RVD sequence: NI-NI-HG-HK-HG-NI-HD-HK-HG-HG-NI (see Figure 7).

[0122] The number of RVDs in the RVD sequence corresponds to the number of repeat units of the Nucleic acid-Binding Domain (NBD). To construct each repeat unit, we select Pair ID: PAIR_014 (RULP: FGPDNLVKVAA and RURP: GGAQALQALLDKGPALLQAG) from Table 13. The sequence of each one repeat unit is presented in Table 28 as follows:

[0123] Table 28: Repeat units of the Nucleic acid-Binding Domain (NBD) for targeting the sequence 5'-AATGTACGTTA-3'

Base	RULP	RVD	RURP	Repeat unit sequence
A	FGPDNLVKVAA	NI	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAG
A	FGPDNLVKVAA	NI	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAG
T	FGPDNLVKVAA	HG	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAG
G	FGPDNLVKVAA	HK	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAG
T	FGPDNLVKVAA	HG	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAG
A	FGPDNLVKVAA	NI	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAG
C	FGPDNLVKVAA	HD	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAG
G	FGPDNLVKVAA	HK	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAG
T	FGPDNLVKVAA	HG	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAG
T	FGPDNLVKVAA	HG	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAG
A	FGPDNLVKVAA	NI	GGAQALQALLDKGPALLQAG	FGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAG

[0124] The Nucleic acid-Binding Domain (NBD) that targets the nucleic acid sequence 5'-AATGTACGTTA-3' is assembled by fusing together, from top to bottom, the repeat unit sequences set forth in Table 28, column 5, row 2-12, giving:

FGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALL
DKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**
GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGP
DNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKG
PALLQAG

[0125] We select MSTAFVDQDKQMANRLNLSPLERSKIEKQYGGATTLAFISNKQNELAQILSRADILKIASYDCAAHALQAVLDCGPMLGKRG (SEQ ID: NTER_005) to be the N-terminal domain. We select RSNEEIVHVAARRGGAGRIRKMVAPLLERQ (SEQ ID: CTER_009) to be the C-terminal domain. The Modular Nucleic Acid-Binding Protein (MNABP) that recognizes the sequence 5'-AATGTACGTTA-3' is obtained by fusing together each part in the following order: N-terminal domain, Nucleic acid-Binding Domain, C-terminal domain, giving:

MSTAFVDQDKQMANRLNLSPLERSKIEKQYGGATTLAFISNKQNELAQILSRADILKIASYDCAAHALQAVLDCGPMLGKRGFGPDNLV
KVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALL
QAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQ
ALLDKGPALLQAGFGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVA

A**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAG
RSNEEIVHVAARRGGAGRIRKMVAPLLERQ

Example 7

[0126] We desire to assemble a Nucleic acid Binding Domain (NBD) that recognizes the nucleic acid sequence 5'-AATGTACGTT**A**-3' (the last nucleotide base is in bold and underlined), which has an RVD sequence of NI-NI-HG-HK-HG-NI-HD-HK-HG-HG-**NI** (the RVD that corresponds to the last nucleotide base is in bold and underlined). We desire to target the last nucleotide base of the target sequence with a half-repeat unit. We select the half-repeat unit SEQ ID: HRU_010 (sequence: FSAADIVKIAS**NN**GGAQALQALIDHWSTLSGKT. The RVD is in bold and underlined) from Table 24 and we substitute its RVD with NI, yielding FSAADIVKIAS**NI**GGAQALQALIDHWSTLSGKT. The sequence of the repeat units that target the first ten nucleotide bases of the target sequence is set forth in herein Table 28, column 5, row 2-11. The sequence of the NBD that targets 5'-AATGTACGTT**A**-3' is thus:
FGPDNLVKVAANIGGAQALQALLDKGPALLQAGFGPDNLVKVAANIGGAQALQALLDKGPALLQAGFGPDNLVKVAAHGGGAQALQALL
DKGPALLQAGFGPDNLVKVAAHKGGAQALQALLDKGPALLQAGFGPDNLVKVAAHGGGAQALQALLDKGPALLQAGFGPDNLVKVAANI
GGAQALQALLDKGPALLQAGFGPDNLVKVAAHDGGAQALQALLDKGPALLQAGFGPDNLVKVAAHKGGAQALQALLDKGPALLQAGFGP
DNLVKVAAHGGGAQALQALLDKGPALLQAGFGPDNLVKVAAHGGGAQALQALLDKGPALLQAG**FSAADIVKIASNIGGAQALQALIDHW
STLSGKT** (the sequence of the half-repeat unit is in bold and underlined).

Example 8

[0127] We desire to construct a Modular Nucleic Acid-Binding Protein (MNABP) comprising: (a) a truncated N-terminal domain derived from the teaching of patent US9902962B2 (Barbas et al., 2012); a Nucleic acid-Binding Domain (NBD) that recognize the sequence 5'- AATGTACGTTA-3'; and (c) a truncated C-terminal domain obtained by deleting the last 202 residues from SEQ ID: CTER_010 (see Table 26), giving
LSPERVAAIACIGGRSAVEAVRQGLPVKAIRRIRREKAPVAGPPPASLGPTPQELVAVLHFFRAHQQPRQAFVDALAAFQATRPALLRLLSS
VGV.

[0128] The truncated N-terminal domain is obtained by deleting the first 127 residues of SEQ ID: NTER_010 (MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDPLRQFDPSLLDTSLFDSMP
AVGTPHTAAAPAEWDEAQSALRAADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQ
HHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITALPEATHEDIVG**VGKQWSGARAL**EALLTKAGELRGPPLQLDTGQLLKIAKRGGV
TAVEAVHASRNALTGAPLN. The first 127 residues are underlined. The sequence VGKQWSGARAL is in bold), and by substituting the W in VGKQ**W**SGARAL with an R (teaching of Barbas et al., 2012), giving
ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQ**R**SGARALEALLTKAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHASRNALTGAPLN. The herein truncated N-terminal domain cause the Modular Nucleic Acid-Binding Protein (MNABP) to recognize preferably a target nucleic acid sequence that begin with a G. We add a 5'-G to the sequence 5'- AATGTACGTTA-3', yielding a target sequence 5'-**G**AATGTACGTTA-3'. We use the sequence of the Nucleic acid-Binding Domain (NBD) set forth in example 6 to target 5'-

AATGTACGTTA-3'. The sequence of the Modular Nucleic Acid-Binding Protein (MNABP) that binds to 5'-**G**AATGTACGTTA-3' is thus:

ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQ**R**SGARALEALLTKAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHASRNALTGAPLNFGPDNLVKVAA**NI**GGAQ
ALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLV
KVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALL
QAGFGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQ
ALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGLSPERVAAIA
CIGGRSAVEAVRQGLPVKAIRRIRREKAPVAGPPPASLGPTPQELVAVLHFFRAHQQPRQAFVDALAAFQATRPALLRLLSSVGV

Example 9

[0129] The goal is the same as in example 8, but the target nucleic acid sequence is 5'-**T**AATGTACGTTA-3'. We will use the same Nucleic-acid Binding Domain (NBC) and C-terminal domain as disclosed in example 8. The truncated N-terminal domain is obtained by deleting the first 127 residues of herein SEQ ID: NTER_010, giving ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQ**W**SGARALEALLTKAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHASRNALTGAPLN. The herein truncated N-terminal domain cause the Modular Nucleic Acid-Binding Protein (MNABP) to recognize preferably a target nucleic acid sequence that begin with a T. The sequence of the Modular Nucleic Acid-Binding Protein (MNABP) that binds to 5'-**T**AATGTACGTTA-3' is thus:

ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQWSGARALEALLTKAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHASRNALTGAPLNFGPDNLVKVAA**NI**GGAQ
ALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLV
KVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALL
QAGFGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQ
ALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGLSPERVAAIA
CIGGRSAVEAVRQGLPVKAIRRIRREKAPVAGPPPASLGPTPQELVAVLHFFRAHQQPRQAFVDALAAFQATRPALLRLLSSVGV

Example 10

[0130] The goal is the same as in example 8, but the target nucleic acid sequence is 5'-**A**AATGTACGTTA-3'. We will use the same Nucleic-acid Binding Domain (NBC) and C-terminal domain as disclosed in example 8. The truncated N-terminal domain is obtained by deleting the first 127 residues of herein SEQ ID: NTER_010, and by substituting the RG in LDTGQLLKIAK**RG**GVTAVEAVHASRNALTGAPLN (RG is in bold) with RSG (teaching of Gregory et al., 2011), giving ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQWSGARALEALLTKAGELRGPPLQLDTGQLLKIAK**RSG**GVTAVEAVHASRNALTGAPLN. The herein truncated N-terminal domain cause the Modular Nucleic Acid-Binding Protein (MNABP) to recognize target nucleic acid sequences with any one of the nucleotide base A, T, G, or C, at their 5' end. The sequence of the Modular Nucleic Acid-Binding Protein (MNABP) that binds to 5'-**A**AATGTACGTTA-3' is thus:

ARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSKHPAALGTVAVTYQHIITA
LPEATHEDIVGVGKQWSGARALEALLTKAGELRGPPLQLDTGQLLKIAK**RSG**GVTAVEAVHASRNALTGAPLNFGPDNLVKVAA**NI**GGA
QALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNL
VKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPAL
LQAGFGPDNLVKVAA**HD**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HK**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQAL
QALLDKGPALLQAGFGPDNLVKVAA**HG**GGAQALQALLDKGPALLQAGFGPDNLVKVAA**NI**GGAQALQALLDKGPALLQAGLSPERVAAI
ACIGGRSAVEAVRQGLPVKAIRRIRREKAPVAGPPPASLGPTPQELVAVLHFFRAHQQPRQAFVDALAAFQATRPALLRLLSSVGV

Example 11

[0131] We desire to construct a Modular Nucleic Acid-Binding Protein (MNABP) that recognizes the target sequence 5'-TAATGGCATGCATCCCA-3'. SEQ ID: NTER_015 is used as the N-terminal domain. SEQ ID: CTER_018 is used as the C-terminal domain. SEQ ID: RU_043 (GGREQVIKIAAH**HG**GQQALQALLDKGPALRQAG) is used as the first repeat unit of the Nucleic acid-Binding Domain (NBD). SEQ ID: RU_005 (FGNDNLVKVAA**NG**GGAQALQALLDKGPALRQAG) is used as the other repeat units. We use the RVDs NI, HG, HK, and HD for the recognition of the nucleotide bases A, T, G, and C, respectively, and the RVDs of SEQ ID: RU_043 and SEQ ID: RU_005 are substituted accordingly. The RVD sequence of the target sequence 5'-TAATGGCATGCATCCCA-3' is HG-NI-NI-HG-HK-HK-HD-NI-HG-HK-HD-NI-HG-HD-HD-HD-NI. The sequence of the Nucleic acid-Binding Domain (NBD) is thus:

GGREQVIKIAA**HG**GGQQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALL
DKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**
GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGN
DNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPALRQAGFGNDNLVKVAAH**DG**GAQALQALLDKG
PALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGA
QALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNL
VKVAA**NI**GGAQALQALLDKGPALRQAG.

[0132] The full sequence of the Modular Nucleic Acid-Binding Protein that binds to 5'-TAATGGCATGCATCCCA-3' is thus:

MNEWIQRHNPQDKQQSSGASVSTQSMVFSQAGSANVSAGVPGPSRTRATHTDTHTVRHSPYPAASARSATSARSANTSSQALSTADHKK
IQKAAGNATLNYVIQHLDELQHALGGREQVIKIAA**HG**GGQQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQA
GFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQAL
LDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**N**
IGGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPALRQAGFG
NDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDK
GPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GG
AQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGVSHDEILALATKQRGASGALQSKLGELTAAGR

[0133] A nuclear localization sequence (sequence: PKKKRKV) is placed at the N-terminus of the herein Modular Nucleic Acid-Binding Protein (MNABP), and the nuclease domain (sequence:

VKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPI

GQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVR
RKFNNGEINF) of the FokI protein is placed at its C-terminus. The resulting Modular Nucleic Acid-Binding Endonuclease (MNABEN) has a sequence:

PKKKRKVMNEWIQRHNPQDKQQSSGASVSTQSMVFSQAGSANVSAGVPGPSRTRATHTDTHTVRHSPYPAASARSATSARSANTSSQAL
STADHKKIQKAAGNATLNYVIQHLDELQHALGGREQVIKIAA**HG**GGQQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDK
GPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GG
AQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDN
LVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HK**GGAQALQALLDKGPA
LRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HG**GGAQA
LQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVK
VAA**HD**GGAQALQALLDKGPALRQAGFGNDNLVKVAA**NI**GGAQALQALLDKGPALRQAGVSHDEILALATKQRGASGALQSKLGELTAAG
RVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSG
GYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGT
LTLEEVRRKFNNGEINF

[0134] A polynucleotide (a DNA sequence) that encodes the herein MNABEN is generated according to mammalian cells codon usage, yielding:

CCCAAGAAGAAGAGGAAGGTGATGAACGAGTGGATCCAGAGGCACAACCCCCAGGACAAGCAGCAGAGCAGCGGCGCCAGCGTTTCTAC
ACAATCTATGGTTTTTTCTCAAGCTGGATCTGCTAATGTTTCTGCTGGAGTGCCTGGACCCTCTAGGACCAGGGCTACCCATACAGACA
CACATACCGTGAGGCATTCTCCTTATCCTGCTGCTTCTGCTAGGTCTGCTACAAGCGCTAGGTCTGCTAACACAAGCAGCCAGGCTTTA
TCTACAGCTGACCACAAGAAGATCCAGAAGGCCGCCGGCAACGCCACCCTGAACTACGTGATCCAACACTTGGACGAGTTACAGCATGC
TCTTGGAGGAAGGGAGCAGGTTATCAAAATTGCCGCTCATGGAGGAGGCCAGCAGGCTTTACAAGCTCTGTTAGATAAGGGACCTGCCT
TGAGGCAGGCCGGCTTCGGCAACGACAACCTGGTTAAAGTTGCTGCTAACATCGGAGGAGCTCAAGCCTTACAGGCTTTATTAGACAAG
GGCCCTGCTCTTAGGCAGGCCGGCTTCGGGAATGATAACCTCGTTAAAGTTGCTGCCAACATCGGCGGAGCTCAAGCCCTTCAAGCTTT
ACTGGATAAAGGACCTGCTCTTAGGCAAGCCGGATTCGGAAACGATAATCTGGTGAAGGTTGCTGCTCATGGAGGCGGAGCTCAGGCCT
TGCAGGCTTTACTGGATAAAGGCCCTGCTTTGAGGCAGGCCGGCTTTGGCAACGACAACTTAGTCAAGGTTGCTGCTCATAAGGGAGGT
GCTCAAGCCTTGCAAGCTTTATTAGACAAGGGACCTGCCCTTAGACAGGCCGGCTTCGGGAACGACAACCTGGTTAAGGTTGCCGCTCA
TAAAGGCGGCGCTCAAGCTCTGCAAGCATTATTAGACAAAGGCCCTGCCTTAAGGCAGGCTGGATTTGGAAACGACAACCTTGTTAAAG
TGGCTGCCCATGACGGAGGCGCTCAGGCCCTTCAGGCACTGCTTGATAAAGGGCCCGCTTTAAGACAGGCCGGCTTTGGAAACGATAAT
CTGGTCAAAGTTGCTGCTAATATAGGAGGCGCCCAAGCCTTACAGGCCTTACTTGATAAGGGACCCGCTCTTAGGCAGGCCGGGTTCGG
CAACGACAATCTTGTGAAAGTTGCTGCCCACGGAGGAGGAGCTCAGGCTTTACAAGCCTTATTAGATAAGGGACCTGCTTTAAGGCAGG
CTGGCTTCGGCAACGACAATCTGGTGAAAGTGGCTGCCCATAAAGGCGGGGCCCAAGCCCTGCAAGCTTTGTTAGATAAAGGTCCCGCC
CTGAGGCAAGCCGGATTCGGTAACGATAATTTAGTTAAAGTCGCTGCTCATGATGGCGGCGCTCAAGCCCTTCAAGCCTTACTGGATAA
GGGACCTGCTCTTAGACAGGCCGGGTTCGGCAATGACAACCTCGTTAAGGTTGCTGCTAATATCGGAGGCGCCCAGGCTTTACAGGCTC
TTTTAGACAAAGGACCTGCTTTAAGGCAAGCCGGCTTCGGCAACGATAACCTTGTGAAAGTTGCTGCTCATGGAGGAGGCGCTCAAGCT
TTGCAGGCTCTGTTGGACAAAGGACCTGCCTTAAGGCAAGCCGGGTTCGGCAATGATAACCTTGTGAAGGTTGCAGCTCATGATGGCGG
AGCCCAAGCTCTTCAGGCCTTGTTAGATAAAGGCCCTGCTCTTAGGCAAGCTGGCTTCGGCAATGACAATCTGGTGAAGGTGGCCGCTC
ATGATGGAGGTGCTCAGGCCCTTCAAGCTTTACTTGATAAAGGCCCCGCTTTAAGGCAGGCCGGCTTCGGAAATGATAACCTGGTTAAA
GTGGCTGCTCATGACGGAGGAGCCCAAGCCTTACAAGCCTTGCTGGACAAAGGCCCTGCTCTTAGACAAGCCGGCTTCGGGAATGACAA
CCTTGTTAAAGTCGCCGCCAACATTGGAGGCGCCCAAGCTTTACAAGCTCTTCTTGACAAAGGACCCGCTTTAAGGCAGGCTGGGGTTA

GCCACGATGAGATCCTGGCTTTAGCTACAAAACAAAGGGGAGCTTCTGGAGCCCTTCAGAGCAAGCTTGGCGAGTTAACAGCCGCTGGC
AGGGTGAAGAGCGAGCTTGAAGAAAAGAAGAGCGAGCTGAGGCATAAGCTGAAGTACGTGCCCCACGAGTACATTGAGCTGATCGAAAT
TGCCAGGAACAGCACCCAGGACAGGATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGCAAACACTTAG
GCGGAAGCAGGAAACCCGATGGCGCCATCTACACAGTGGGATCTCCCATTGATTATGGCGTGATCGTGGACACCAAGGCCTACAGCGGA
GGCTATAATTTACCCATTGGCCAAGCTGACGAGATGCAGAGGTACGTTGAGGAGAACCAGACCAGGAACAAGCACATCAACCCCAACGA
GTGGTGGAAGGTGTACCCCTCTAGCGTTACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAATTACAAGGCCCAGCTGA
CAAGGCTGAACCACATCACCAACTGCAACGGAGCTGTGTTGAGCGTTGAAGAGCTGCTGATCGGAGGAGAGATGATCAAGGCCGGCACA
TTGACCCTTGAAGAAGTTAGGAGGAAGTTCAACAACGGCGAGATCAACTTC. A restriction site (*PmeI*, sequence: GTTTAAAC) and a Kozak consensus sequence (sequence: **GCCACC<u>ATG</u>G**CC, the consensus is in bold, the initiation codon is underlined) is placed upstream of the herein polynucleotide, while a STOP codon (sequence: TAG) and a restriction site (*NotI,* sequence: GCGGCCGC) is placed downstream of it.

[0135] The polynucleotide sequence is chemically synthesized and cloned into the pcDNA3.1(+) vector (its Multiple Cloning Site (MCS) is in the forward (+) orientation) by applying a standard cloning protocol (restriction enzymes *PmeI* and *NotI* are used). Standard bacterial transformation and plasmid purification protocols are applied to obtain sufficient amount of the herein Modular Nucleic Acid-Binding Endonuclease (MNABEN) encoding plasmid (pMNABEN).

[0136] pMNABEN is transfected into mammalian cells (which harbor MNABEN binding sites) using a standard transfection protocol. The gene encoding MNABEN is expressed under a strong and constitutive CMV promoter. Biosynthesized MNABEN polypeptides bind to two sequences 5'-TAATGGCATGCATCCCA-3' (disposed in forward and reverse orientation) in such a way that their endonuclease FokI domains dimerize (see Figure 8) and mediate a double-strand break.

REFERENCES

Barbas CF, Mercer A, Lamb BM, et al. 2012. Chimeric polypeptides having targeted binding specificity. Patent US9902962B2

Bertonati C, Duchateau P, Juillerat A, et al. 2012. New modular base-specific nucleic acid binding domains from burkholderia rhizoxinica proteins. Patent WO2014018601A2

Boch J, Scholze H, Schornack S, et al. 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science 326: 1509–1512

Bonas U, Boch J, Schornack S, et al. 2009. Modular DNA-binding domains and methods of use. Patent US9017967B2

Cunnac S, Occhialini A, Barberis P, et al. 2004. Inventory and functional analysis of the large Hrp regulon in Ralstonia solanacearum: identification of novel effector proteins translocated to plant host cells through the type III secretion system. Molecular Microbiology 53: 115–128

de Lange O, Schreiber T, Schandry N, et al. 2013. Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. New phytologist 199(3):773-86

de Lange O, Wolf C, Dietze J, et al. 2014. Programmable DNA-binding proteins from Burkholderia provide a fresh perspective on the TALE-like repeat domain. Nucleic Acids Research. 42(11):7436-49.

Gregory PD, Miller JC, Paschon D, et al. 2011. Modified dna-binding proteins and uses thereof. Patent EP2780460B1

Ralf KÜ, Wurst W. 2014. RALEN-mediated genetic modification techniques. Patent WO2014167058A1

Schornack S, Meyer A, Römer P, et al. 2006. Gene-for-gene mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. Journal of Plant Physiology 163: 256–272

Urnov F, Stamatoyannopoulos JA, Funnell AP. 2018. Animal pathogen-derived polypeptides and uses thereof for genetic engineering. Patent WO2019204643A2

Voytas D, Bogdanove AJ, Zhang F, et al. 2010. Tal effector-mediated DNA modification. Patent WO2011072246A2

Zhang F, Le Cong, Kosuri S, et al. 2011. Transcription activator-like effectors. Patent US9499592B2